



Published in final edited form as:

J Proteome Res. 2008 April ; 7(4): 1693–1703. doi:10.1021/pr700706s.

Plasma Glycoprotein Profiling for Colorectal Cancer Biomarker Identification by Lectin Glycoarray and Lectin Blot

Yinghua Qiu[†], Tasneem H. Patwa[†], Li Xu^{‡, #}, Kerby Shedden[§], David E. Misek^{‡, Δ}, Missy Tuck^Δ, Gracie Jin[‡], Mack T. Ruffin[⊥], Danielle K. Turgeon[∇], Sapna Synal[◊], Robert Bresalier[¶], Norman Marcon⁺, Dean E. Brenner^{Δ, ∇}, and David M. Lubman^{*, †, ‡, Δ}

[†]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

[‡]Department of Surgery, University of Michigan Medical Center, Ann Arbor, Michigan 48109

[#]Cancer Center, Sun Yat-sen University, Guang Zhou, PRC, 510060

[§]Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109

^ΔComprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109

[⊥]Department of Family Medicine, University of Michigan, Ann Arbor, Michigan 48109

[∇]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109

[◊]Dana-Farber Cancer Institute/Brigham and Woman's Medical Center, Boston, Massachusetts 02115

[¶]Department of Gastroenterology, Hepatology & Nutrition, University of Texas MD Anderson Cancer Center, Houston, Texas 77030

⁺Department of Gastroenterology, St. Michael's Hospital, University of Toronto, Toronto, Canada

Abstract

Colorectal cancer (CRC) remains a major worldwide cause of cancer-related morbidity and mortality largely due to the insidious onset of the disease. The current clinical procedures utilized for disease diagnosis are invasive, unpleasant, and inconvenient; hence, the need for simple blood tests that could be used for the early detection of CRC. In this work, we have developed methods for glycoproteomics analysis to identify plasma markers with utility to assist in the detection of colorectal cancer (CRC). Following immunodepletion of the most abundant plasma proteins, the plasma N-linked glycoproteins were enriched using lectin affinity chromatography and subsequently further separated by nonporous silica reverse-phase (NPS-RP)-HPLC. Individual RP-HPLC fractions were printed on nitrocellulose coated slides which were then probed with lectins to determine glycan patterns in plasma samples from 9 normal, 5 adenoma, and 6 colorectal cancer patients. Statistical tools, including principal components analysis, hierarchical clustering, and Z-statistic analysis, were employed to identify distinctive glycosylation patterns. Patients diagnosed with colorectal cancer or adenomas were shown to have dramatically higher levels of sialylation and fucosylation as compared to normal controls. Plasma glycoproteins with aberrant glycosylation were identified by nano-LC-MS/MS, while a lectin blotting methodology was used to validate proteins with significantly altered glycosylation as a function of cancer progression. The potential markers identified in this study for diagnosis to distinguish colorectal cancer from adenoma and normal include elevated sialylation and fucosylation in complement C3, histidine-rich glycoprotein, and kininogen-1. These potential

markers of colorectal cancer were subsequently validated by lectin blotting in an independent set of plasma samples obtained from 10 CRC patients, 10 patients with adenomas, and 10 normal subjects. These results demonstrate the utility of this strategy for the identification of N-linked glycan patterns as potential markers of CRC in human plasma, and may have the utility to distinguish different disease states.

Keywords

Plasma glycoproteomics; lectin affinity enrichment; lectin glycoarrays; lectin blot; nano-LC-MS/MS

1. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide with an estimated one million new cases and a half-million deaths each year, largely due to the insidious onset of the disease.¹ Although colorectal cancer accounts for 14% of all cancer-related deaths, the overall 5-year survival rate could be around 90% if the cancer was diagnosed and treated early, while the tumors were still localized.² The most widely used screening procedures for CRC include fecal occult blood testing, flexible sigmoidoscopy, double-contrast barium enema, and colonoscopy.^{3–5} However, as these screening methods are not pleasant, there is patient resistance to undergo the procedure. Thus, serum/plasma-based screening methodologies for CRC detection are likely to be far more acceptable than the current screening options and would likely greatly increase the percentage of the population screened.

There is currently great interest in proteomics-based plasma biomarkers with utility for the early detection of cancer, as well as to follow the efficacy of therapeutic intervention.⁶ The plasma proteome may contain not only normal plasma proteins, but also the different cleavage products and proteins with different post-translational modifications that may result from a given disease state.⁷ Within the plasma proteome, the plasma or serum glycoproteome is both functionally important and perhaps the most abundant post-translational modification.^{8–23} Many existing cancer biomarkers are glycoproteins, such as Her2/neu in breast cancer, prostate-specific antigen (PSA) in prostate cancer, CA125 in ovarian cancer, and carcinoembryonic antigen (CEA) in colorectal, bladder, breast, pancreatic, and lung cancer. Although CEA remains the most widely used serum glycoprotein biomarker in CRC, poor sensitivity and specificity precludes its use for the early detection of CRC.²⁴ Other tumor antigens that have been proposed as serum glycoprotein biomarkers of colorectal cancer include CA 19-9, CA 242, CA-195, CA 50, CA 74-2, and TIMP-1 (tissue inhibitor of metalloproteinase-1).^{25–32} However, as with CEA, these biomarkers have poor sensitivity and specificity, and are not suitable for screening or diagnostic purposes.

Alterations in either the level of or type of glycosylation has been shown to influence cellular processes such as growth, differentiation, transformation, adhesion, metastasis, and immune surveillance of tumors.^{33–35} In particular, aberrant sialylation in cancer cells is thought to be a characteristic feature associated with malignant properties including invasiveness and metastatic potential. An increase in sialylation is commonly observed in various tumors, which may be due to either an increased activity of sialyltransferases or increased numbers of possible sialylation sites on N-linked carbohydrates.³⁶ Recent studies in prostate cancer have also shown changes in fucosylation associated with progression of the disease.^{37,38} Thus, it is of great interest to identify and validate plasma glycoproteins with altered glycosylation whose function may reveal insight into critical events in cancer progression and may have utility as potential markers for cancer detection. In this study, we utilized multidimensional liquid chromatography protein separation followed by lectin glycoarrays for screening N-

glycosylation pattern changes in plasma from patients with colorectal cancer. Bioinformatics analysis of the data facilitated the identification of plasma glycoproteins that possess altered glycosylation. These potential biomarkers were subsequently validated with a second set of independent plasma samples. These glycoprotein biomarkers may have utility for the detection of colorectal cancer.

2. Materials and Methods

2.1. Plasma Samples

Human plasma samples were collected through a four institutional consortium (Dana Farber Cancer Institute, MD Anderson Cancer Center, St. Michael Hospital, Toronto, Ontario, Canada, and University of Michigan Medical Center) of the Early Detection Research Network (EDRN). Human subjects were identified prior to endoscopy and informed consent obtained prior to sample collection procedures specified in a protocol approved by Institutional Review Boards at all collaborating Institutions. The samples were collected, handled, shipped, stored, and managed according to standard operating procedures as specified in the protocol document. Samples were labeled with bar-coded subject identification number and tracked from collection through assay via a relational database containing deidentified demographic and clinical data located at the Bioinformatics Unit at Dartmouth Medical College. The samples were stored in a professional repository facility at -80°C until use. The plasma was obtained from 6 patients with colorectal cancer (two stage II, two stage III, and two stage IV), 5 samples from patients with colonic adenomas (polyp size with 0.3–1.3 cm), and 9 samples from patients with normal colonoscopes for use in a blinded set to screen N-linked glycosylation pattern changes on plasma glycoproteins and 30 plasma samples (10 of each) for use in a testing set. All subjects that donated plasma for this study were between 50 and 76 years of age with 16 Caucasians and 4 African Americans. The plasma was aliquoted into 0.5 mL aliquots, frozen, and then stored at -80°C until assayed.

2.2. Preparation of Glycoprotein Samples for Lectin Glyco-arrays or Lectin Blot

2.2.1. Delipidation and Immunodepletion of the Plasma Samples—The plasma samples were delipidated by centrifugation for 15 min at 20 000g, and the lipid-containing upper layer was removed before depletion. A total of 250 μL of the delipidated plasma was depleted using the ProteomeLab IgY-12 LC10 proteome partitioning kit (Beckman Coulter, Fullerton, CA). This procedure enables simultaneous removal of 12 highly abundant proteins from human plasma, including albumin, IgG, α 1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α 1-acid glycoprotein, α 2-macroglobin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen. Using optimized buffers for sample loading, washing, eluting, and regeneration, the resulting flow-through (unbound) fraction and the eluted (bound) fraction were collected separately during a total of 75 min IgY affinity separation cycle. The final depleted fraction was buffer-exchanged into 2mL Concanavalin A (Con A) binding buffer (20 mM Tris, 0.15 M NaCl, 1 mM Mn^{2+} , and 1 mM Ca^{2+} , pH 7.4) with a 10 000 Da molecular weight limit Amicon Ultra-15 centrifugal filter (Millipore, Billerica, MA). The protein concentration of the final concentrated fraction was determined using a Bradford protein assay (Bio-Rad, Hercules, CA) with BSA as a standard. The concentrations of the immunodepleted plasma samples were approximately 1.5–2.0 mg/mL.

2.2.2. N-Glycoprotein Enrichment with ConA Affinity Capture—ConA columns were prepared by adding 1.5 mL of the agarose-bound lectin (Vector Laboratories, Burlingame, CA) into 5 mL polypropylene columns (Pierce Biotechnology, Rockford, IL). The columns were first equilibrated with 5 mL of the binding buffer before use. A total of 500 μL of the immunodepleted plasma was loaded onto an equilibrated column. After incubating for 30 min, the unbound proteins were washed out with 6 mL of the binding buffer, and the captured

proteins were eluted with 4 mL of the elution buffer (20 mM Tris, 0.5 M NaCl, and 0.5 M methyl-*R*- β -mannopyranoside, pH 7.4). The protein recovery of the lectin column was determined based on the Bradford protein assay, using BSA as the standard.

2.2.3. HPLC Separation of Glycoproteins—Twenty-five micrograms of the enriched *N*-glycoproteins (corresponding to around 60 μ L of original plasma) was separated by NPS-RP-HPLC at a flow rate of 0.5 mL/min on a 33 \times 4.6 mm ODS III column (Eprogen, Darien, IL) using a ProteomeLab PF2D system (Beckman Coulter, Fullerton, CA). The separation was performed using a water (solvent A) and acetonitrile (solvent B) gradient as follows: (1) 5–25% B in 1 min; (2) 25–31% B in 2 min; (3) 31–37% B in 7 min; (4) 37–41% B in 8 min; (5) 41–48% B in 7 min; (6) 48–58% B in 3 min; (7) 58–75% B in 1 min; (8) 75–100% B in 1 min. Proteins eluted from the column were collected by an automated fraction collector (FC 204 BE, Beckman-Coulter) controlled by an in-house-designed DOS-based software program and 32 Karat software (Beckman-Coulter). The 32 Karat software was also used to calculate the peak area of each protein fraction.

2.3. Lectin Glycoarrays

After completely drying, the protein fractions were resuspended with 15 μ L of printing buffer (65 mM Tris-HCl, 1% SDS, 5% DTT, and 1% glycerol) in 96 well plates, and then arrayed on nitrocellulose slides (Whatman, Keene, NH) using a noncontact piezoelectric printing device (Nanoplotter 2.0, GeSiM, Germany). Volumes of 2.5 nL of each fraction were arrayed on the nitrocellulose slides in spots that were 450 μ m in diameter and 600 μ m apart. The printed slides were dried for 1 day after being blocked overnight with 1% BSA in phosphate buffered saline with 0.1% Tween 20 (PBS-T). The blocked slides were first incubated with biotinylated lectin for 2 h and then with 1 μ g/mL streptavidin conjugated to Alexaflor555 fluorescent dye (Invitrogen, Carlsbad, CA). After being washed and dried, slides were scanned in the green channel using an Axon 4000A scanner. Image analysis was performed using the GenePix 6.0 software (Molecular Devices, Sunnyvale, CA).

2.4. Statistical Analysis of Lectin Glycoarray Data

2.4.1. Principal Components Analysis (PCA)—Principal components analysis (PCA) was performed for data visualization, which was carried out using log-transformed and normalized array spot intensities. The leading two eigenvectors of the sample covariance matrix were used for visualization. In this study, 20 plasma samples (processed in duplicate when using ConA, AAL, and PNA and triplicate when using SNA and MAL) were placed in a two-dimensional scatter plot using PCA. Sample pairs falling close together in the scatter plot are more similar in terms of their overall patterns of normalized glycol-form abundances. The PCA was based on all microarray measurements without selection or weighting. All samples were included in the analysis without selection.

2.4.2. Hierarchical Clustering—An unsupervised hierarchical clustering (HC) procedure was used without any prior knowledge of grouping to find criteria appropriate for classifying the cases according to the glycosylation pattern from glycoarrays. To do this, the normalized array spot intensities were log-transformed, and the pairwise Pearson correlations were used to carry out HC in which more closely correlated pairs of samples were joined at a lower point on the dendrogram. The scale on the dendrograms was 100–100 \times *r*, where *r* is the Pearson correlation coefficient. In the HC analysis, the replicate averages of the 20 distinct biological specimens were used.

2.4.3. Z-Statistics—For differential abundance analysis, Z-statistics for each protein detected by each lectin were calculated. The Z-statistic is the difference in mean levels between two groups being compared (based on log₂ data) divided by an estimate of its standard error.

For single comparisons, Z-statistics greater than approximately 2 in magnitude correspond to *p*-values smaller than 0.05. The Z-statistics of differentially glycosylated proteins detected by lectins together with fold changes both in log 2 and non-log 2 forms are shown in Supplementary Table 1 in Supporting Information. Comparisons were made of normal versus adenoma and normal versus cancer, as well as adenoma versus cancer. On the basis of the Bonferroni correction for two-sided testing of 36 peaks, Z-values of ≥ 3.2 or ≤ -3.2 could be deemed to have significantly different glycosylation levels at a 95% significance level.

2.5. SDS-PAGE and Lectin Blot

To identify and validate the glycoproteins of interest as markers of CRC, protein fractions from NPS-RP-HPLC were further separated by 1-D SDS-PAGE using the Mini-Protean cell (Bio-Rad, Hercules, CA) at 80 V. The resolved proteins were stained with colloidal Coomassie (Invitrogen) or transferred onto a polyvinylidene fluoride (PVDF) membrane (Bio-Rad). The PVDF membrane was blocked with 5% (w/v) BSA (Roche, Indianapolis, IN) in PBS-T overnight at 4 °C and then incubated with biotinylated AAL and SNA, respectively (2 µg/mL in PBS-T containing 3% BSA) for 1 h at room temperature. The membrane was then washed and incubated with a 100 ng/mL streptavidin-HRP in PBS-T containing 3% BSA. After washing, the signal was visualized using a chemiluminescence detection system (ECL, Pierce) and detected on blue sensitive autoradiography film (Marsh Bio Products, Rochester, NY). Potential glycoprotein biomarkers are detected by lectin blot experiment, and the corresponding colloidal blue stained bands were identified by nano-LC-MS/MS.

2.6. Protein Digestion

2.6.1. Tryptic Digestion and N-Deglycosylation of NPS-RP-HPLC Fractions—The NPS-RP-HPLC fractions with significantly different glycosylation were dried completely, denatured in 40 µL of 100 mM NH₄HCO₃ buffer (pH 7.8), then reduced with 1 mM dithiothreitol (DTT) for 45 min at 56 °C and alkylized with 15 mM iodoacetamide (IAA) for 1 h at room temperature in the dark. The proteins were then digested with 1–2 µg of TPCK-treated trypsin (Promega, Madison, WI) for 18 h at 37 °C. The reaction mixture was then heated for 10 min at 95 °C to stop the action of trypsin. A total of 1–2 µL of PNGase F (New England BioLabs, Ipswich, MA) was added to half of the tryptic digest mixture from each fraction to start the N-deglycosylation at 37 °C for 12 h. The other half was stored at –80° for later use.

2.6.2. Tryptic Digestion and N-Deglycosylation of SDS Gel Bands—The glycoprotein bands from the colloidal Coomassie blue stained SDS-PAGE gel were carefully excised. The gel pieces were placed in siliconized Eppendorf tubes (Sigma), destained three times with 200 µL of 200 mM ammonium bicarbonate and 40% acetonitrile at 37 °C for 30 min each, and lyophilized completely in a SpeedVac (Thermo). The dried gel pieces were first deglycosylated by incubating with 10 µL of the PNGase F solution (Sigma) overnight at 37 °C followed by trypsin digestion overnight at 37 °C. The liquid from the gel piece was transferred to a new tube for nano-LC-MS/MS analysis.

2.7. Mass Spectrometry for Protein Identification and Glycosylation Site Determination

A MS4B MDLC system (Michrom Bioresources, Auburn, CA) interfaced with a linear ion trap mass spectrometer (LTQ, Thermo, San Jose, CA) was used to analyze the tryptic digests from SDS-PAGE. The injected peptide sample was first desalted on a trap column (150 µm × 50 mm, Michrom Bioresources Inc., Auburn, CA) with 3% solvent B (0.3% formic acid in acetonitrile) at 50 µL/min for 5 min and then released and separated on a nano column (150

Supporting Information Available: Table of Z-statistics of differentially glycosylated proteins detected by lectins. This material is available free of charge via the Internet at <http://pubs.acs.org>.

$\mu\text{m} \times 150 \text{ mm}$, Michrom) using a 45 min gradient from 3% B to 95% B at 0.3 $\mu\text{L}/\text{min}$. The resolved peptides were directly introduced into an ESI ion source with the spray voltage set at 2.6 kV (Figure 1).

To identify the eluted peptides, data-dependent MS/MS analysis (m/z 400–2000) was performed using MS acquisition software (Xcalibur 1.4, Thermo Finnigan), in which a full MS scan was followed by seven MS/MS scans of the seven most intense precursor ions. All MS/MS spectra were compared against the Swiss-Prot FASTA human protein database using the SEQUEST algorithm incorporated into the TurboSequest feature of Bioworks 3.1 SR1.4 (Thermo Finnigan). When up to two missed cleavages were allowed, positive protein identification was accepted for a peptide with X_{corr} of greater than or equal to 3.5 for triply-, 2.5 for doubly-, and 1.9 for singly charged ions, and all with $\Delta\text{Cn} \geq 0.1$.³⁹ The sequence database search was set to accept the following modifications: carboxymethylated cysteines due to treatment with iodoacetamide, oxidized methionines, and an enzyme-catalyzed conversion of asparagines to aspartic acids (0.984 Da shift) at an N-glycosylation site. Accuracy of the SEQUEST assignment of MS/MS spectra to peptide sequences was validated by the TransProteomics Pipeline which includes both PeptideProphet and ProteinProphet software. In this study, peptides were identified with a probability cutoff of $p \geq 0.99$, and protein identifications were confirmed with probability scores of at least 0.9.

3. Results and Discussion

3.1. Reduction in the Complexity of Plasma Glycoprotein Mixtures by Immunoaffinity Depletion and Lectin Affinity Enrichment

The strategy undertaken during analysis of protein glycosylation will vary depending on the amount of available sample. It is more challenging to determine alterations in glycan structure when the glycoproteins are present in a complex biological medium at a very wide dynamic range, such as in human serum or plasma. To analyze glycoproteins expressed in the midlevel abundance range, the most abundant proteins were immunodepleted from the sample, as shown in the schematic flowchart of the proposed method (Figure 1). A total of 250 μL of each plasma sample was first delipidated and then immunodepleted to remove the lipids and the 12 most abundant plasma proteins based on an avian antibody (IgY)- antigen interaction. Around 7% of the total protein mass in the plasma samples remained after the immunodepletion step. The representative chromatograms resulting from the immunodepletion of plasma from normal, adenoma, and colorectal cancer patients (Figure 2A) indicate the reproducibility of this step.

Following immunodepletion, the remaining proteins in the flow-through fraction were subjected to ConA affinity chromatography to enrich the concentration of N-glycosylated proteins. With broad specificity and high affinity, ConA binds with preference to oligomannosidic, hybrid, and biantennary N-glycans, either unconjugated or attached to proteins or peptides.⁴⁰ O-Glycopeptides or glycoproteins that contain exclusively O-glycosylation sites are not bound by this lectin. Approximately 70% of the immunodepleted plasma protein content was recovered by ConA affinity chromatography. By selectively isolating the glycoproteins from the immunodepleted plasma proteome, the procedure achieved a significant reduction in analyte complexity at two levels: first, the immunodepletion of the 12 most abundant proteins significantly increases the dynamic range of detection by approximately 90- fold and, additionally, reduces sample heterogeneity due to the removal of the highly variable IgG, IgA, and IgM proteins; second, the subsequent N-glycoprotein enrichment step affords another effective means of reducing plasma sample complexity.

Twenty-five micrograms of the enriched N-glycoprotein mixtures were further separated by NPS-RP-HPLC into 36 fractions for lectin glycoarray or lectin blot analysis. Figure 2B shows the reverse-phase chromatograms of the plasma samples from different physiological status (9

normals, 5 adenomas, and 6 colorectal cancers). The reproducibility of these chromatograms indicates that the samples from the plasma from normal subjects, from adenoma patients, and from colorectal cancer patients were very similar at the protein expression level. These results suggest that the analysis of glycoprotein expression alone does not provide valuable information to differentiate the clinical status of individuals.

3.2. Lectin Glycoarrays for Identification of N-Glycosylation Pattern Changes

To analyze the plasma glycosylation patterns, all fractions containing the separated intact glycoproteins from the NPS-RP-HPLC separation were then arrayed on nitrocellulose slides, in duplicate, as unique spots. Subsequently, the slides were screened to analyze the different glycan structures using five different lectins. The following biotinylated lectins were used: *Aleuria aurantia* lectin (AAL), *Sambucus nigra* bark lectin (SNA), *Maackia amurensis* lectin II (MAL), peanut agglutinin (PNA), and Concanavalin A (ConA). AAL binds fucose linked (α -1,6) to *N*-acetylglucosamine or (α -1,3) to *N*-acetylglucosamine related structure. Both MAL and SNA recognize sialic acid on the terminal branches. MAL detects glycans containing NeuAc-Gal-GlcNAc with sialic acid at the 3 position of galactose, while SNA binds preferentially to sialic acid attached to terminal galactose in an (α -2,6) and an (α -2,3) linkage at a lesser degree. In contrast, PNA binds desialylated exposed galactosyl (β -1,3) *N*-acetylgalactosamine. ConA recognizes α -linked mannose including high-mannose-type and hybrid-type structures. The utilization of these five lectins have been proved to be highly successful in covering >95% of N-glycan types reported and differentiating them according to their specific structures.⁴¹ The array was analyzed using five different lectins for comparing N-glycosylation levels in plasma from normal, adenoma, and colorectal cancer patients. Since only variations in glycan expression were of interest, all array spot intensities were normalized by dividing the corresponding UV peak area to eliminate protein abundance differences. The normalized array data suggest that the overall levels of protein fucosylation and sialylation are higher in colorectal cancer and adenoma plasma samples as compared to the normal plasma controls.

3.3. Statistical Analysis of N-Glycosylation Pattern Changes

Principal components analysis (PCA) and hierarchical clustering (HC) of the normalized glycoarray data were performed to differentiate the plasma samples in terms of their overall N-glycosylation patterns and to relate these patterns to clinical status. In PCA, 20 plasma samples assayed in duplicate (ConA, AAL, PNA) or triplicate (MAL, SNA) were analyzed separately for each lectin. The scores of the first two principal components of the normal, adenoma, and colorectal cancer samples are illustrated in a 2-dimensional scatter plot in which each sample was plotted as an individual point (Figure 3A). Closer points corresponded to greater similarity in the patterns of glycoprotein expression over all 36 protein spots on the microarrays. When ConA and SNA were hybridized against the arrays, the normal controls (red) were grouped separately from cancer (blue) and adenoma samples (green), while most cancer and adenoma samples were clustered together. In response to AAL and MAL, the normal and cancer samples generally segregated from each other, whereas the adenoma samples overlapped to some extent with both. The PNA-based microarrays did not provide robust fluorescence intensities for most protein spots; however broadly similar results compared to AAL and MAL arrays were observed. Except for PNA, all the other lectins clearly separate the cancer samples from normal controls. The results of the PCA analysis suggest that lectin glycoarrays may have utility as a diagnostic tool to discriminate the diseased states from the nondiseased states in cancer detection. The excellent concordance among the replicates from the same sample (Figure 3B) in PCA results indicates that the lectin glycoarray is a robust strategy for screening N-glycosylation changes among the plasma samples from different disease states. As expected, similar results were observed in HC by using the Pearson correlation coefficient for distance metrics (Figure 4). The clustering results for fucosylated,

sialylated, and mannosylated glycan expression generally distinguished the normal plasma samples from the cancer and adenoma samples. The results from the different lectins indicate the effectiveness of using multilection detection to differentiate plasma samples of the different clinical states based on N-glycosylation pattern changes.

As an alternative means of analyzing the lectin glycoarray data, we calculated Z-statistics of each array spot to search for signature proteins that might differentiate the plasma samples of the different clinical states (Table 1). Comparisons were performed of normal versus adenoma (N/A), normal versus cancer (N/C), and adenoma versus cancer (A/C). Z-values of ≥ 3.2 or ≤ -3.2 were selected as differential glycosylation at a 95% significance level. A positive Z-value indicates elevated glycosylation and a negative Z-value means reduced glycosylation.

3.4. Identification of Plasma Biomarkers with Altered N-Glycosylation

Because of the coelution during the NPS-RP-HPLC separation, there were cases where more than one protein was observed in certain fractions. To determine which coeluting protein was responsible for the differential responses in the lectin-based microarrays, the fraction with altered glycosylation was further separated by 1-D SDS-PAGE and then analyzed by lectin blot. Since the elevated fucosylation and sialylation levels in colorectal cancer plasma were detected on most of the differentially glycosylated proteins, we chose AAL and SNA in the lectin blot analysis to determine which protein corresponded to the differential fucosylation and sialylation pattern.

The corresponding protein bands in the SDS gel with significantly differential glycosylation patterns in colorectal cancer or adenoma were excised, and then digested with PNGase F and trypsin. Protein identification and the possible glycosylation sites were determined by nano-LC-MS/MS. Positive identification was validated by the Trans-Proteomics pipeline which includes PeptideProphet and ProteinProphet software. PeptideProphet software was used to effectively identify correct peptide assignments, and ProteinProphet was used to validate the protein identifications. In this study, peptides were identified with probability scores of at least 0.99 with a false-positive error rate of 0.0007, and proteins were identified with a probability cutoff of $p \geq 0.9$ which corresponds to a 0.7% error rate.^{42,43} Figure 5A shows a representative nano-LC-ESI-MS/MS spectrum of a deglycosylated glycopeptide $[(M + 2H)^{2+}$ at m/z 553.20] from complement C4. The localization of the N-glycosylation site was determined by a mass increase of 1 Da on the N-X-(S/T) sequence after deamidation of asparagine residue into aspartic acid.⁴⁴ The b- and y- series of product ions clearly showed a mass shift indicative of conversion of asparagine to aspartic acid at the original site of N-glycosylation. In this case, the mass difference of 115 Da for aspartic acid found for both the b_3 - b_2 and y_9 - y_8 product ion pairs suggests the original N-glycosylation at residue 3. Figure 5B shows another example of a peptide $[(M + 2H)^{2+}$ at m/z 716.82] from kininogen-1. Again a shift of 115 Da for both b_6 - b_5 and y_7 - y_6 indicates the N-glycosylation at residue 6, while b_5 - b_4 and y_8 - y_7 yield a difference of 114 Da indicating that the Asn at position 5 was not N-glycosylated. The significant differentially glycosylated proteins with their Z-statistics are summarized in Table 1, and the corresponding detected glycosylation sites are shown in Table 2, in which 10 proteins displayed significant differential glycosylation among normal, adenoma, and cancer samples. Three of these proteins showed elevated glycosylation in the case of cancer compared to normal and adenoma, and seven had higher glycosylation levels in cancer and adenoma compared to normal. Many of the proteins identified may not be specific to colon cancer but are due to systemic changes and may be acute phase proteins or proteins from the liver or pancreas. Nevertheless, the data suggests that Z-statistic analysis of lectin glycoarrays has utility to identify cancer samples relative to adenoma or normal controls. The potential markers to distinguish colorectal cancer from adenoma and normal controls identified in this study include

the elevated sialylation and fucosylation in complement C3, histidine-rich glycoprotein, and kininogen-1.

3.5. Lectin Blot of a Control Set for Detection of Potential Biomarkers for Differentiating the Different Clinical States

The correlation between the potential biomarkers and a diagnosis of colorectal cancer or adenoma has been confirmed in a blinded sample set with 30 plasma samples (10 colorectal cancers, 10 adenomas, and 10 normal). The plasma samples were depleted, enriched and separated by multidimensional HPLC separation as described previously and then analyzed by 1-D SDS-PAGE, followed by lectin blotting. Figure 6 shows representative protein bands after lectin blot which characterized the CRC samples. As shown, complement C3 in all of the normal and adenoma samples barely responded in either the AAL or SNA blot, but was observed to have significantly elevated fucosylation and sialylation in the colorectal cancer samples. These results from lectin blot analysis were consistent with that obtained from the Z-statistic analysis in which complement C3 in cancer was significantly elevated in response to both AAL and SNA as compared to adenoma and normal. The peak areas of complement C3 in each plasma sample indicated that approximately equal amounts of protein were loaded. Additionally, histidine-rich protein displayed significant differential glycosylation but similar protein expression. In this case, fucosylation was found to be significantly elevated in colorectal cancer samples as compared to both adenoma and normal samples, while similar sialylation was observed in cancer and adenoma. Again, the total amounts of histidine-rich protein were quite similar among the plasma samples from normal, adenoma, and cancer patients. These results highlight the potential utility of the altered glycosylation patterns instead of absolute protein expression as markers for cancer detection and further increase the specificity of these potential markers.

4. Conclusion

We have described a glycoproteomic strategy for the identification of potential plasma biomarkers in the detection of colorectal cancer. The strategy was based on the reduction of plasma complexity by immunodepletion and subsequent glycoprotein enrichment, the screening of glycan pattern changes by the lectin glycoarray format, and the identification of potential markers with altered glycosylation using lectin blot analysis and nano-LC-MS/MS. As indicated by the peak intensities from NPS-RP-HPLC separations, the absolute protein amounts of each clinical state were quite similar so that the plasma glycoproteome alone does not differentiate the clinical status of individuals. When lectin glycoarrays were used, normal, adenoma, and colorectal cancer plasma showed distinct clustering results of each state following the PCA and HC analysis. In this study, patients diagnosed with either colorectal cancer or adenomas have dramatically higher levels of sialylation and fucosylation compared to the normal controls. The glycoprotein fractions for analysis were identified using SDS-PAGE and lectin blot coupled with nano-LC-MS/MS. In future work, further analysis of sialic acid and fucose containing glycans will be done with more advanced mass spectrometry techniques.⁴⁵ The potential markers identified in this study to distinguish colorectal cancer from adenoma and normal include elevated sialylation and fucosylation in complement C3, histidine-rich glycoprotein, and kininogen-1. These results demonstrated the usefulness of this strategy for the identification of the N-linked glycan patterns to distinguish individuals in different clinical states as well as the identification of potential biomarkers of CRC in human plasma based upon changes in glycan structure rather than in the protein level.

Abbreviations

ConA, Concanavalin A; AAL, Aleuria aurentia lectin; MAL, *Maackia amurensis* lectin; PNA, peanut agglutinin; SNA, *Sambucus nigra* bark lectin; NPS-RP-HPLC, nonporous silica reverse-phase high-performance liquid chromatography.

Acknowledgment

This work was supported by the National Cancer Institute under grants R01CA106402 and R21CA124441 and by the NCI-EDRN under CA86400.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. CA Cancer J. Clin 2005;55:74–108. [PubMed: 15761078]
2. Engwegen JY, Helgason HH, Cats A, Harris N. World J. Gastroenterol 2006;12:1536–1544. [PubMed: 16570345]
3. Burt RW. Gastroenterology 2000;119:837–853. [PubMed: 10982778]
4. Ransohoff DF. Gastroenterology 2005;128:1685–1695. [PubMed: 15887159]
5. Kung JW, Levine MS, Glick SN, Lakhani P. Radiology 2006;240:725–735. [PubMed: 16837671]
6. Wulfkuhle JD, Liotta LA, Petricoin EF. Nat. Rev. Cancer 2003;3:267–275. [PubMed: 12671665]
7. Feldman AL, Espina V, Petricoin EF, Liotta LA, Rosenblatt KP. Surgery 2004;135:243–247. [PubMed: 14976472]
8. Novotny MV, Mechref Y. J. Sep. Sci 2005;28:1956–1968. [PubMed: 16276785]
9. Plavina T, Wakshull E, Hancock WS, Hincapie M. J. Proteome Res 2007;6:662–671. [PubMed: 17269723]
10. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS. Clin. Chem 2006;52:1897–1905. [PubMed: 16916992]
11. Qiu R, Zhang X, Regnier FE. J. Chromatogr., B 2007;845:143–150.
12. Qiu R, Regnier FE. Anal. Chem 2005;77:7225–7231. [PubMed: 16285669]
13. Madera M, Mechref Y, Klouckova I, Novotny MV. J. Proteome Res 2006;5:2348–2363. [PubMed: 16944947]
14. Zhou Y, Aebersold R, Zhang H. Anal. Chem 2007;79:5826–5837. [PubMed: 17591751]
15. Zhang H, Liu AY, Loriaux P, Wollscheid B. Mol. Cell. Proteomics 2007;6:64–71. [PubMed: 17030953]
16. Haab BB, Geierstanger BH, Michailidis G, Vitzthum F. Proteomics 2005;5:3278–3291. [PubMed: 16038022]
17. Block TM, Comunale MA, Lowman M, Steel LF. Proc. Natl. Acad. Sci. U.S.A 2005;102:779–784. [PubMed: 15642945]
18. Steel LF, Shumpert D, Trotter M, Seeholzer SH. Proteomics 2003;3:601–609. [PubMed: 12748940]
19. Drake RR, Schwegler EE, Malik G, Diaz J. Mol. Cell. Proteomics 2006;5:1957–1967. [PubMed: 16760258]
20. Shin S, Cazares L, Schneider H, Mitchell S. J. Am. Coll. Surg 2007;204:1065–1071. [PubMed: 17481542]
21. Ueda K, Katagiri T, Shimada T, Irie S. J. Proteome Res 2007;6:3475–3483. [PubMed: 17705522]
22. Hu S, Loo JA, Wong DT. Proteomics 2006;6:6326–6353. [PubMed: 17083142]
23. Zhang H, Chan DW. Cancer Epidemiol. Biomarkers Prev 2007;16:1915–1917. [PubMed: 17932335]
24. Fletcher RH. Ann. Int. Med 1986;104:66–73. [PubMed: 3510056]
25. Duffy MJ. Ann. Clin. Biochem 1998;35:364–370. [PubMed: 9635101]
26. Kuusela P, Haglund C, Roberts PJ. Br. J. Cancer 1991;63:636–640. [PubMed: 2021550]
27. Ward U, Primrose JN, Finan PJ, Perren TJ. Br. J. Cancer 1993;67:1132–1135. [PubMed: 8494712]
28. Eskelinen M, Pasanen P, Kulju A, Janatuinen E. Anticancer Res 1994;14:1427–1432. [PubMed: 8067718]

29. Lindmark G, Kressner U, Bergström R, Glimelius B. *Anticancer Res* 1996;16:895–898. [PubMed: 8687147]
30. Carpelan-Holmström M, Haglund C, Lundin J, Alfthan H. *Br. J. Cancer* 1996;74:925–929. [PubMed: 8826859]
31. von Kleist S, Hesse Y, Kananeeh H. *Anticancer Res* 1996;16:2325–2331. [PubMed: 8694564]
32. Holten-Andersen MN, Christensen IJ, Nielsen HJ, Stephens RW. *Clin. Cancer Res* 2002;8:156–164. [PubMed: 11801553]
33. Hakomori S. *Adv. Exp. Med. Biol* 2001;491:369–402. [PubMed: 14533809]
34. Hakomori S. *Proc. Natl. Acad. Sci. U.S.A* 2002;99:225–232. [PubMed: 11773621]
35. Choudhury A, Moniaux N, Ulrich AB, Schmied BM. *Br. J. Cancer* 2004;90:657–664. [PubMed: 14760381]
36. Orntoft TF, Vestergaard EM. *Electrophoresis* 1999;20:362–371. [PubMed: 10197444]
37. Barrabés S, Pagès-Pons L, Radcliffe CM, Tabarés G. *Glycobiology* 2007;17:388–400. [PubMed: 17229815]
38. Zhao J, Qiu W, Simeone DM, Lubman DM. *J. Proteome Res* 2007;6:1126–1138. [PubMed: 17249709]
39. Qian W, Liu T, Monroe ME, Strittmatter EF. *J. Proteome Res* 2005;4:53–62. [PubMed: 15707357]
40. Cummings RD, Kornfeld S. *J. Biol. Chem* 1982;257:11230–11234. [PubMed: 7118880]
41. Patwa TH, Zhao J, Anderson MA, Simeone DM, Lubman DM. *Anal. Chem* 2006;64:11–6421. [PubMed: 16970316]
42. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. *Anal. Chem* 2002;74:5383–5392. [PubMed: 12403597]
43. Yan W, Lee H, Deutsch EW, Lazaro CA. *Mol. Cell. Proteomics* 2004;3:1039–1042. [PubMed: 15269248]
44. Gonzalez J, Takao T, Hori H, Besada V. *Anal. Biochem* 1992;205:151–158. [PubMed: 1443554]
45. Mazsaroff I, Yu W, Kelley BD, Vath JE. *Anal. Chem* 1997;69:2517–2524. [PubMed: 9212710]

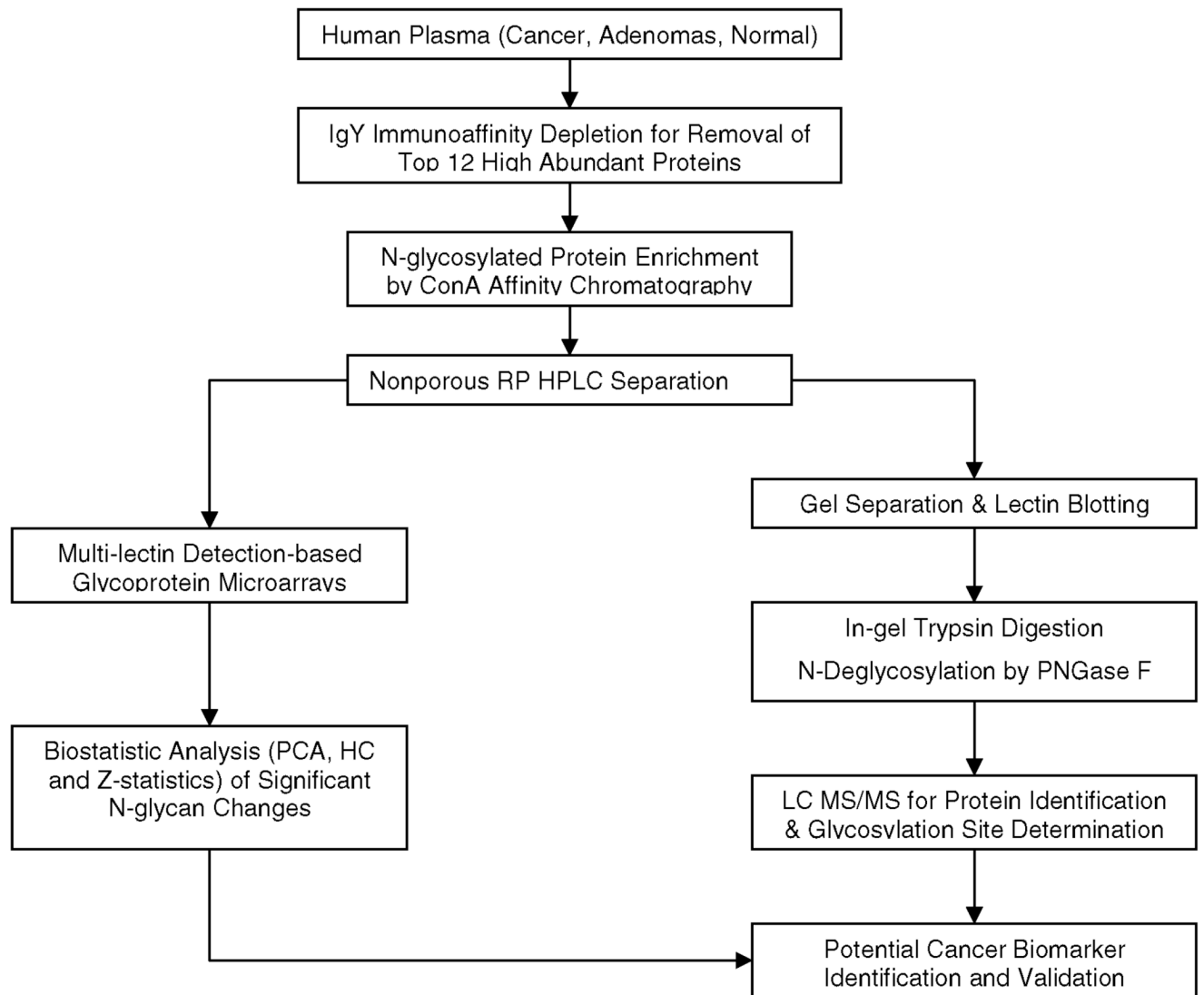


Figure 1. Schematic presentation for high-throughput analysis of plasma N-glycosylation pattern changes in colorectal cancer.

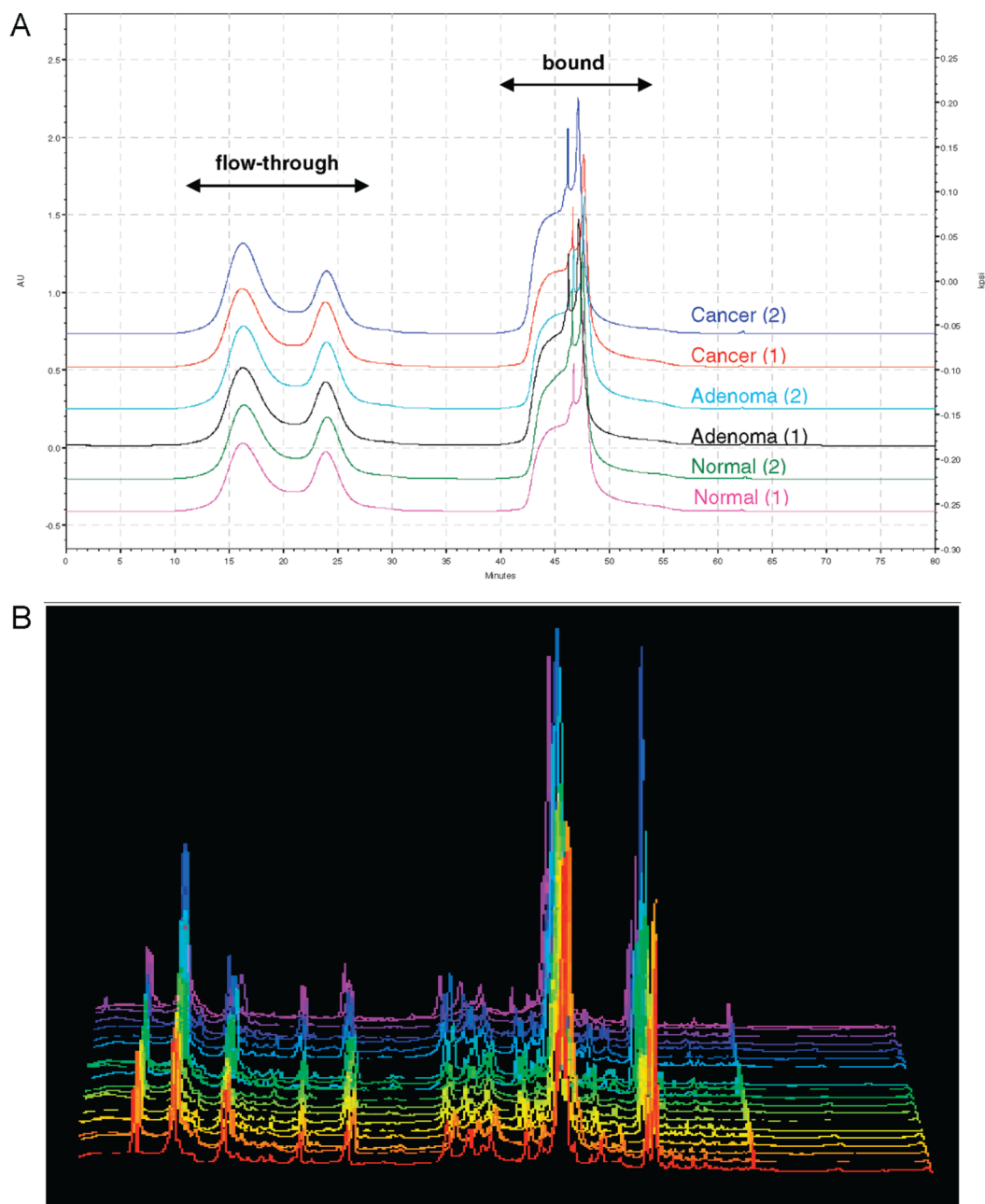


Figure 2.

(A) Representative chromatographic profiles of immunoaffinity depletion of plasma from normal, adenoma, and colorectal cancer patients using ProteomeLab IgY-12 kit. The 12 most abundant proteins are contained in the “bound” fraction, while the less abundant proteins in plasma or serum remained in the “flow-through” fraction. (B) UV chromatograms of all the plasma samples from colorectal cancer, adenoma, and normal controls. The similarity among these UV chromatograms among different samples indicated that proteins undergo heterogeneity of glycosylation modifications without necessarily changing protein expression.

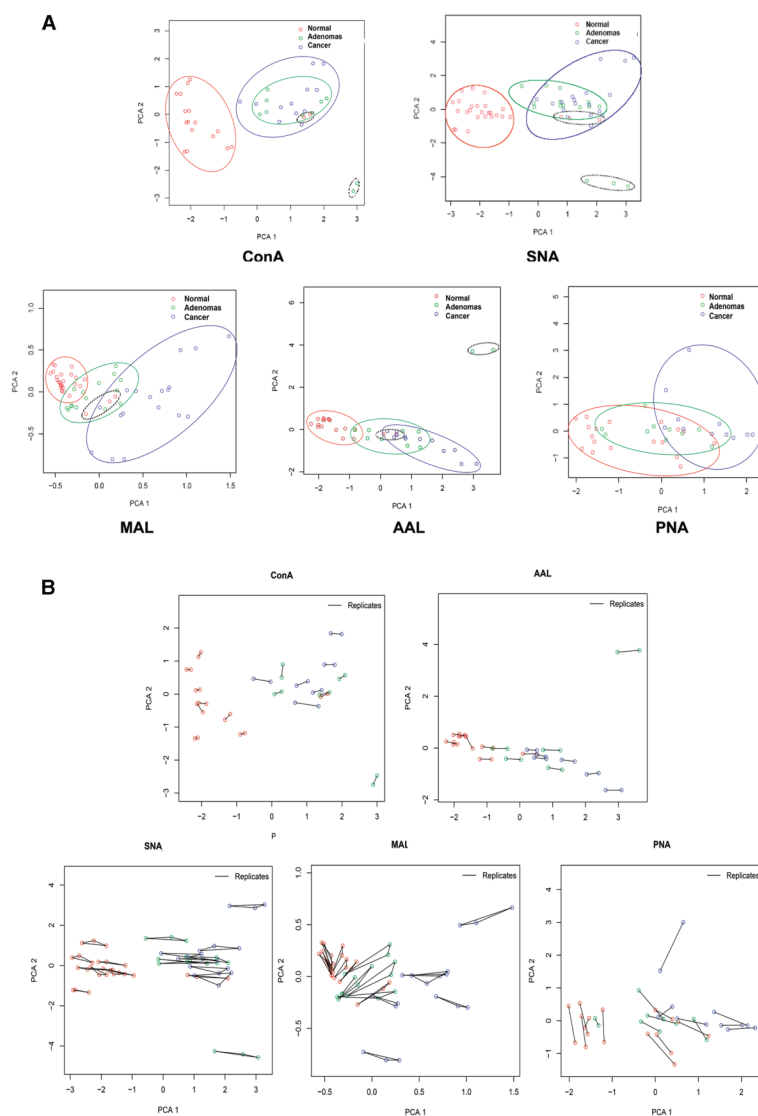


Figure 3. (A) Principal component analysis (PCA) plot for normalized glycoprotein microarray data derived from the replicates of healthy individuals, adenoma, and colorectal cancer patients. Ovals indicate the areas where the data points of the three groups are distributed. (B) Reproducibility demonstration of Principal components analysis (PCA) for normalized glycoprotein microarray data derived from the replicates of healthy individuals, adenoma, and colorectal cancer patients.

Plasma Glycoprotein Profiling for CRC Biomarker Identification

research articles

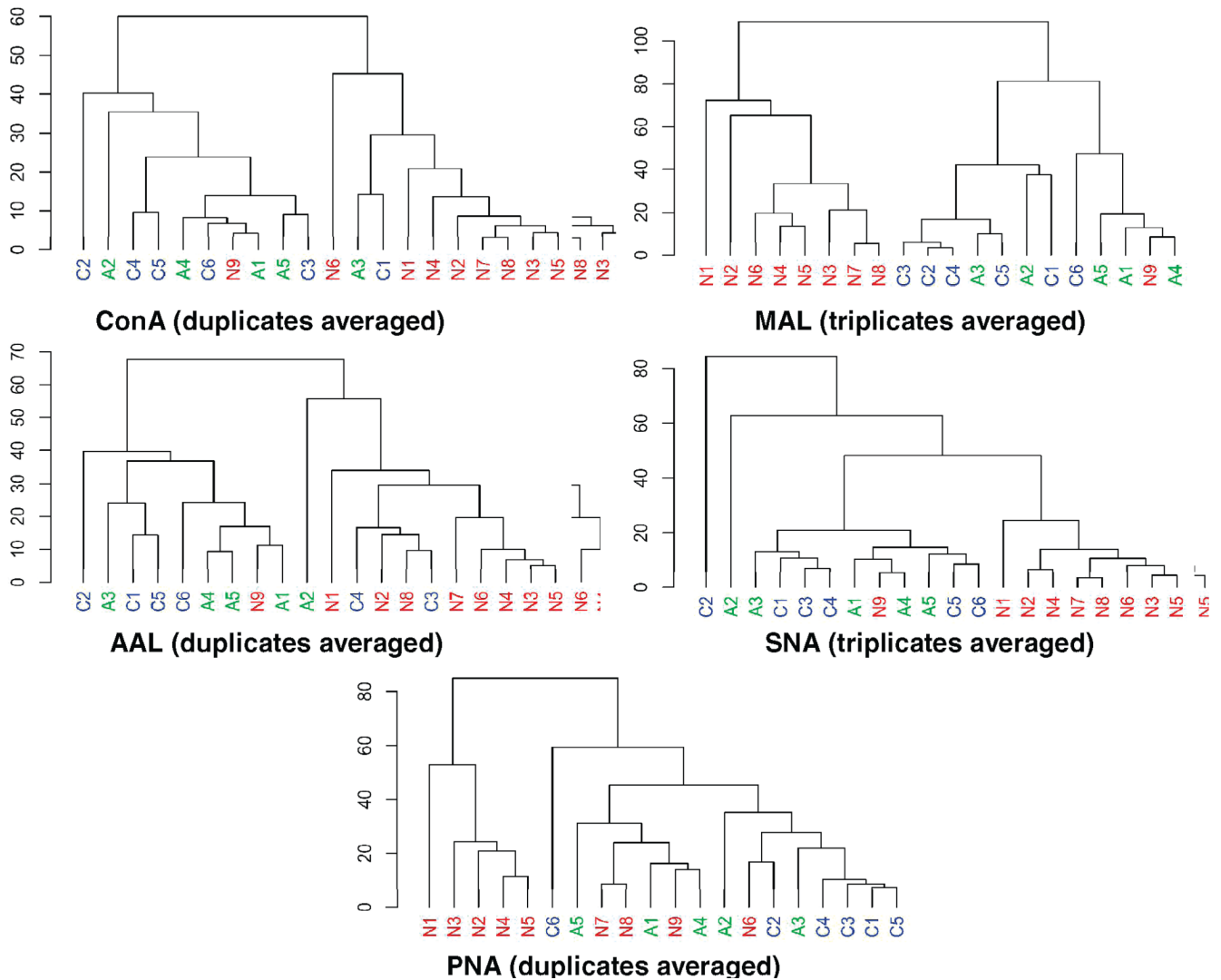


Figure 4. Unsupervised hierarchical clustering of glycoprotein microarray data distinguishes colorectal cancer (c1–c6) from adenoma (a1–a5) and normal controls (n1–n9). The clustering method was the average linkage, and the dissimilarity was obtained from the Pearson correlation coefficient.

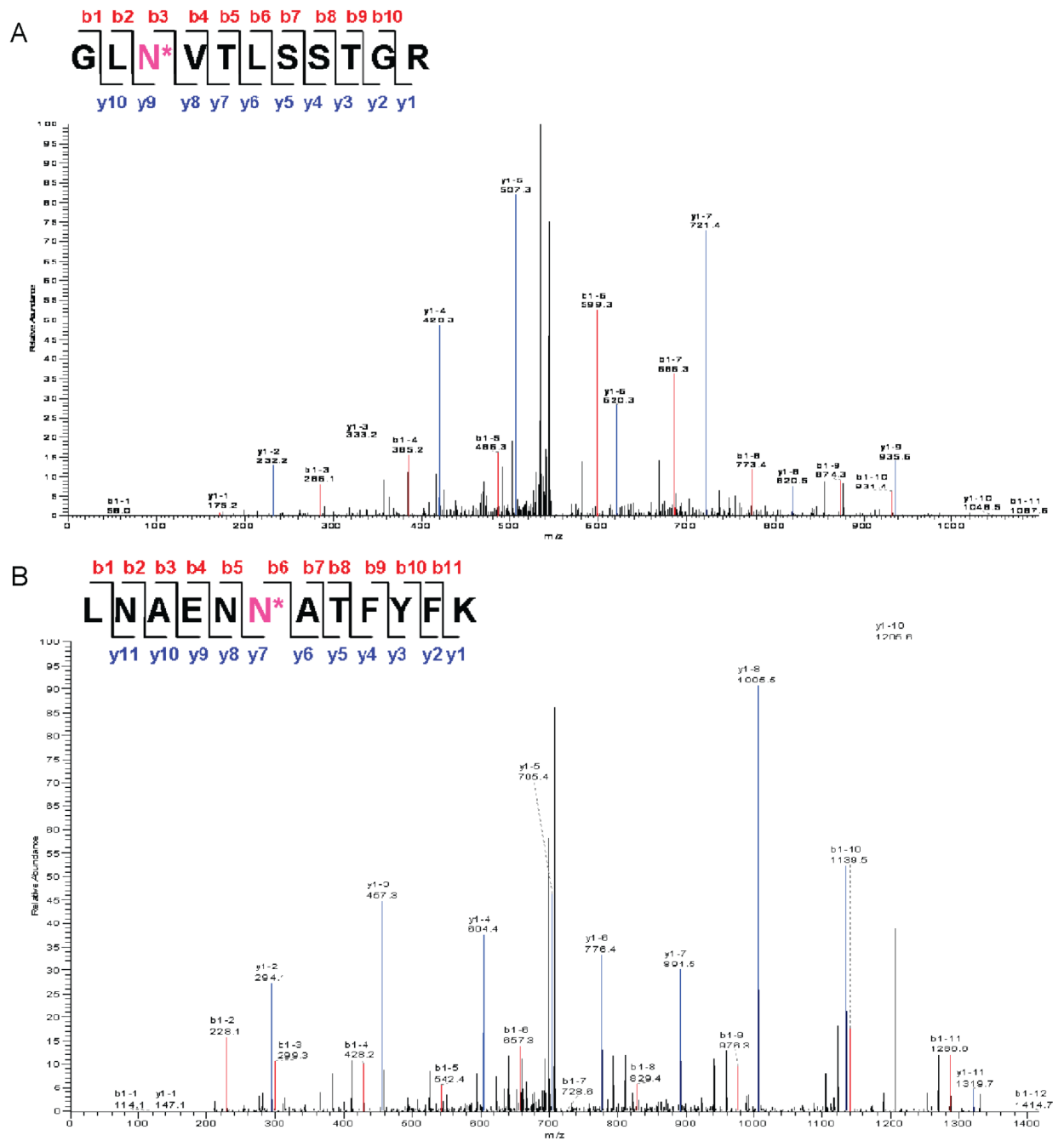


Figure 5. Nano-LC-MS/MS mass spectra of (A) doubly charged N-glycosylated peptide GLN*VTLSSGH ($m/z = 553.28$) from complement 4 and (B) doubly charged N-glycosylated peptide LANEN*ATFYFK from kininogen-1. The asterisk (*) denotes the site of N-glycosylation as determined from the tandem mass spectrum.

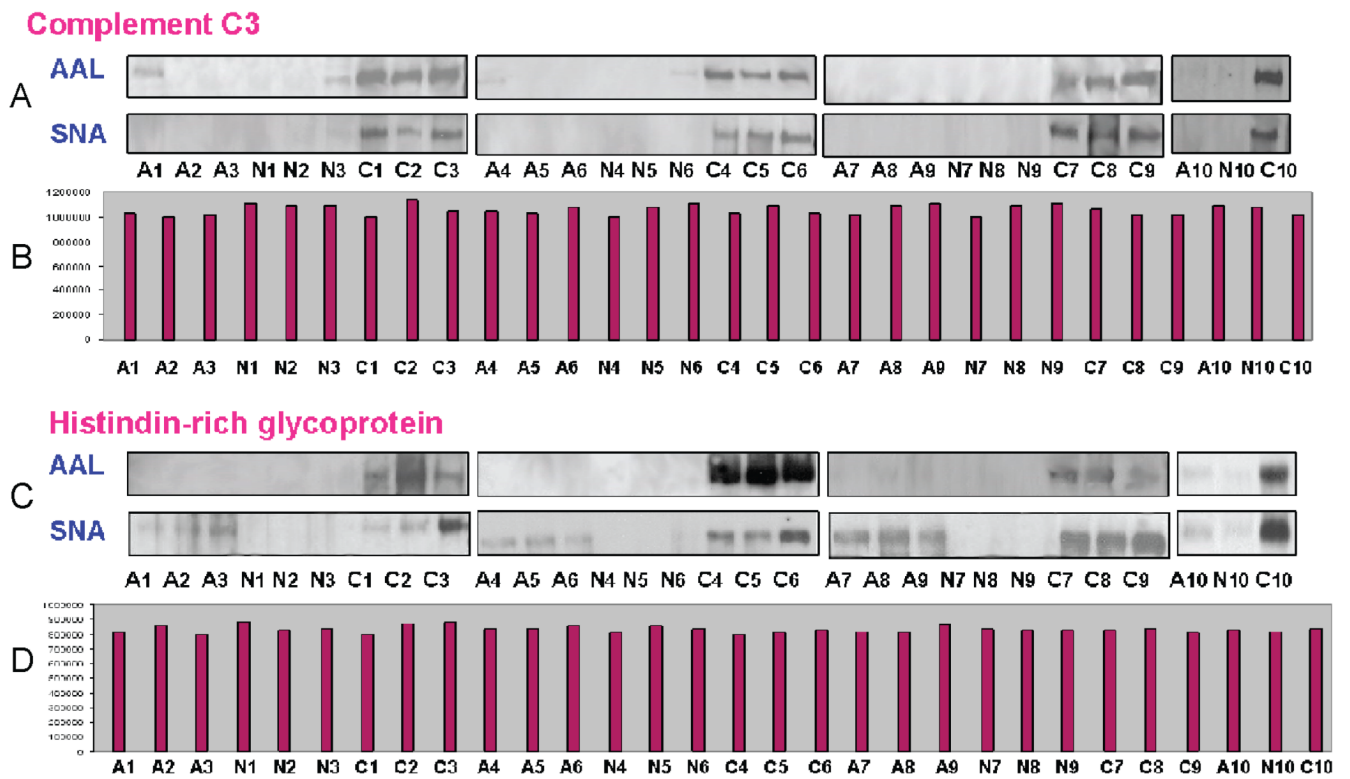


Figure 6. Elevated fucosylation and sialylation of complement C3 (A) and histidine-rich glycoprotein (C) investigated by AAL and SNA blot analysis. The corresponding protein expression levels are shown in (B) for complement C3 and (D) for histidine-rich glycoprotein, respectively.

Table 1

Z-Statistics of Differentially Glycosylated Proteins Detected by Lectins^a

protein ID (access number)	ConA			AAL			MAL			SNA			PNA		
	N/A	N/C	A/C	N/A	N/C	A/C	N/A	N/C	A/C	N/A	N/C	A/C	N/A	N/C	A/C
Complement C3 (P01024)	0.6	-0.09	-1.93	-1.9	-4.19	-3.35	-2.91	-5.38	-4.05	-1.72	-5.77	-3.35	-1.51	-6.21	-4.61
Kininogen-1 (P01042)	-4.86	-6.48	0.01	-5.04	-7.22	-3.44	-2.73	-7.64	-4.75	-6.68	-10.0	-3.24	-1.26	-4.67	-2.94
Histidine-rich glycoprotein (P04196)	-1.25	-2.23	-0.53	-0.55	-4.03	-3.64	0.75	-1.89	-2.86	-1.37	-3.33	-2.44	0.84	-0.98	-2.52
Alpha-1B-glycoprotein (P04217)	-3.29	-3.94	-0.75	-3.04	-6.94	-1.43	-1.65	-2.65	-0.93	-3.24	-5.13	-1.65	0.04	-5.17	-4.69
Hemopexin (P02790)	-6.41	-5.86	1.32	-6.68	-5.77	0.18	-2.95	-3.03	-0.12	-7.62	-7.01	0.58	-1.28	-2.57	-0.8
Complement factor I (P05156)	-2.57	-3.89	-0.52	-2.32	-3.28	-1.07	-0.98	-2.09	-1.11	-3.48	-5.60	-1.15	-0.44	-4.28	-2.81
Ceruloplasmin (P00450)	-4.61	-4.30	0.50	-4.06	-4.57	-0.94	-3.00	-2.52	-0.3	-5.06	-6.65	0.24	-0.01	-4.02	-3.61
Afamin (P43652)	-4.47	-4.35	0.82	-3.86	-4.80	-2.04	-0.29	-2.11	-2.09	-4.19	-4.34	-0.74	-2.38	-1.64	-1.30
Alpha-1-antichymotrypsin (P01011)	-4.21	-5.96	0.89	-3.14	-5.32	0.27	-1.13	-1.05	0.47	-4.07	-5.82	1.08	-1.34	0.48	1.68
Complement C4 precursor (P01028)	-3.80	-5.95	0.07	-3.42	-6.05	0.69	-2.56	-2.34	1.59	-4.22	-6.09	0.95	-1.88	-2.25	0.41

^aN, normal; A, adenoma; C, cancer. The numbers highlighted in bold ($Z \geq 3.2$ or $Z \leq -3.2$) correspond to 95% significant level with multiple testing correction.

Table 2
Differentially Glycosylated Proteins Identified with the Glycosylation Sites

protein ID (access no.)	MW/pI	Peptide Sequence	glycosylation site	MH+
Histidine-rich glycoprotein (P04196)	59541.9/7.09	R.VIDFNC#TTSSVSSALANTK.D	125	2017.96
		R.HSHNNNSDDLHPHK.H	344	1623.74
Kininogen-1 (P01042)	71901.1/6.34	K.LNAENNATFYFK.I	294	1431.69
Hemopexin (P02790)	51644.3/6.55	R.SWPAVGNC#SSALR.W	187	1347.65
		K.ALQPQNVTSLLGC#TH.-	453	1678.86
Complement factor I (P05156)	65677.6/7.72	K.FLNNGTC#TAEGK.F	103	1254.58
Alpha-1B-glycoprotein (P04217)	54239.6/5.58	R.EGDHEFLEVPEAQEDVEATFPVHQPGNYSCSYR.T	179	3779.65
Ceruloplasmin (P00450)	122128.6/5.44	K.AGLQAFFVQEC#NK.S	358	1595.69
		K.EHEGAIYPDNTTDFQR.A	138	1892.84
Afamin (P43652)	69025.0/5.64	R.YAEDKFNETTEK.S	402	1474.67
		R.DIENFNSTQK.F	33	1195.56
Alpha-1-antichymotrypsin (P01011)	47621.5/5.33	K.YTGNASALFILPDQDK.M	271	1752.88
Complement C3 (P01024)	187046.9/6.02	K.TVLTPATNHMGNVFTTIPANR.E	85	2260.08
		K.HYLMWGLSSDFWGEKPNLSYIIGK.D	1617	2841.41
Complement C4 (P01028)	192651.5/6.66	R.FSDGLESNSSTQFEVK.K	226	1774.81
		R.GLNVTLSSSTGR.N	1328	1104.60