

RAPID COMMUNICATION

Lectin precipitation using phytohemagglutinin-L₄ coupled to avidin–agarose for serological biomarker discovery in colorectal cancer

Yong-Sam Kim^{1*}, Ok Lye Son^{1*}, Ju Yeon Lee², Sun Hee Kim¹, Sejeong Oh³, Yoon Suk Lee³, Cheorl-Ho Kim⁴, Jong Shin Yoo², Jeong-Hwa Lee⁵, Eiji Miyoshi⁶, Naoyuki Taniguchi⁶, Samir M. Hanash⁷, Hyang Sook Yoo¹ and Jeong Heon Ko¹

¹ Daejeon-KRIBB-FHCRC Research Cooperation Center, KRIBB, 111 Gwahangno, Yuseong-gu, Daejeon, Korea

² Division of Proteome Research, Korea Basic Science Institute, Yuseong-gu, Daejeon, Korea

³ Department of Surgery, Our Lady of Mercy Hospital, The Catholic University of Korea, Incheon, Korea

⁴ Department of Biological Sciences, Sungkyunkwan University, Suwon City, Kyunggi-Do, Korea

⁵ Department of Biochemistry, College of Medicine, The Catholic University of Korea, Banpo-Dong, Seoul, Korea

⁶ Department of Biochemistry, Osaka University Medical School/Graduate School of Medicine, Suita, Japan

⁷ Fred Hutchinson Cancer Research Center, Seattle, WA, USA

N-acetylglucosaminyltransferase V (GnT-V) has been reported to be upregulated in malignant cancer cells, and its targets have been sought after with regard to biomarker identification. The low capacity and high false positive rates of 2-DE gel-based lectin blots using phytohemagglutinin-L₄ (L-PHA) prompted us to develop a novel protocol for identifying GnT-V targets, in which serum proteins were subjected to immunodepletion, alkylation, and lectin precipitation using L-PHA coupled to avidin–agarose bead complexes, and tryptic digestion. Proteins captured by L-PHA conjugates were analyzed by a nano-LC-FT-ICR/LTQ MS. Here, we report 26 candidate biomarkers for colorectal cancer (CRC) that show 100% specificity and sensitivities of greater than 50%. Not only can these candidate proteins be used as analytes for validation, but the novel protocol described herein can be applied to biomarker discovery in nonCRCs.

Received: January 14, 2008

Revised: March 24, 2008

Accepted: March 28, 2008

Keywords:

Biomarker / Colorectal cancer / GnT-V / L-PHA

Nearly 800 000 new CRC cases are believed to develop each year globally, which accounts for approximately 10% of all incident cancers. In addition, the mortality from CRC is estimated to reach nearly 450 000 deaths annually [1]. The MLH1 and MSH genes are associated with hereditary non-polyposis CRC [2], and the APC gene is associated with

familial adenomatous polyposis [3], but these factors fail to account for the occurrence of a wide range of CRC. Moreover, CRC is just one of many epithelium-derived cancers in which circumstantial factors govern over hereditary genetic factors. These attributes necessitate a clear marker that serves as a tracer molecule for the efficacious treatment of CRC, but unfortunately, a discrete biomarker for CRC has yet to be discovered.

Traditional approaches have stressed the dynamics of protein expression levels associated with the biochemical processes of cancer. In contrast, many lines of evidence have demonstrated the role of various glycosyltransferases in the pathogenicity of cancer cells, wherein changes in protein glycosylation have been reported to be associated with the

Correspondence: Dr. Jeong Heon Ko, Daejeon-KRIBB-FHCRC Research Cooperation Center, KRIBB, 111 Gwahangno, Yuseong-gu, Daejeon 305-806, Korea

E-mail: jhko@kribb.re.kr

Fax: +82-42-879-8119

Abbreviations: AFP, α -fetoprotein; CRC, colorectal cancer; GlcNAc, *N*-acetylglucosamine; GnT-V, *N*-acetylglucosaminyltransferase V; ID, immunodepletion; L-PHA, phytohemagglutinin-L₄

* Both authors contributed equally to this work.

pathogenic processes of cells [4]. One of the best-characterized glycosyltransferases is *N*-acetylglucosaminyltransferase V (GnT-V), which catalyzes the formation of a β 1,6-*N*-acetylglucosamine (GlcNAc) moiety in an N-linked core glycan. An increase in β 1,6-branching on N-linked glycans is associated with the metastatic potential of cancer cells [5]. Several target molecules for GnT-V have been proposed to be involved in cancer progression, including matriptase [6], β 1 integrin [7], N-cadherin [8], and TIMP-1 [9]. Although these proteins were found to be relevant to the biological and pathological processes in cancers, none of them has demonstrated any use as a serum biomarker, which may be due to the fact that molecular fluctuations observed at the cellular level are buffered by the complexity of blood composition, or that current proteomics technologies are unable to trace changes in minute expression level of the proteins in blood.

Despite the strong implication of GnT-V in cancer pathology, no targets of GnT-V have been identified in serum as biomarker candidates. Thus, in this report, we attempted to identify colorectal cancer (CRC)-related GnT-V targets in serum using phytohemagglutinin-L₄ (L-PHA), a lectin that recognizes β 1,6-GlcNAc moiety attached to an N-linked core glycan. Though L-PHA binds rather specifically to β 1,6-GlcNAc moiety of N-glycans, a method of detachment for L-PHA-based lectin affinity chromatography is not available, which has limited use of the lectin to the gel-based lectin blot analysis for search for GnT-V targets. However, the gel-based approach innately has several drawbacks: (i) the 2-DE gel has limited protein-loading capacity, and (ii) despite the development of cutting-edge, high-sensitivity mass spectrometers, a ostensibly single protein spot can contain several scores of proteins, in which case selection of a true positive is often difficult. These two drawbacks make it even more difficult to mine candidate proteins for GnT-V by lectin blot analysis in serum that shows high dynamic range of proteins. In this regard, we developed a simple and reliable protocol wherein GnT-V targets are captured, precipitated, tryptic-digested, and identified using an FT-ICR-LTQ mass spectrometer. Here, we report our optimized protocol, as well as 26 candidate biomarkers for CRC discovered through this method.

Serum samples were prepared from CRC patients at Our Lady of Mercy Hospital at The Catholic University of Korea (Inchon, Korea) and from healthy volunteers at KRIBB (Daejeon, Korea), with an agreement of participation from all subjects. Serum samples were diluted two to three-fold with an appropriate buffer and subjected to partial purification performed with a ProteomPrep 20 plasma immunodepletion (ID) LC column kit (Sigma) or by Con A-affinity column chromatography (GE Healthcare). With a ProteomPrep 20 plasma ID LC column (Sigma), 20 highly abundant serum proteins were removed according to the manufacturer's instructions. For Con A-affinity column chromatography, diluted sera were loaded onto a Con A-agarose column pre-equilibrated with 50 mM Tris-HCl (pH 7.4); the column was then washed with equilibration buffer, and bound proteins were eluted with 0.4 M β -D-methylmanno pyranoside in

50 mM Tris-HCl (pH 7.4). The partially purified samples were reduced by treatment with 1% v/v β -mercaptoethanol and alkylated with excess iodoacetamide at room temperature for 1 h. The modified samples were desalted on a HiPrep 26/10 desalting column (GE Healthcare) and concentrated to a final volume of 1 mL. Protein samples were precleared with avidin-agarose beads for 1 h at room temperature, and the precleared proteins were allowed to bind to L-PHA-avidin-agarose or L-PHA-agarose complexes overnight at 4°C. After extensive washing with PBS, the bound proteins were separated from the bead complexes by adding 1 \times SDS-PAGE denaturation buffer or 6 M urea. Protein preparations that were denatured by addition of denaturation buffer were in-gel digested as described previously [10]. Protein preparations denatured in 6 M urea were diluted to 0.6 M for tryptic digestion. The digested peptides were lyophilized in a SpeedVac system and analyzed by MS.

The peptide mixtures were loaded onto a C18 trap column (5 μ m, 100 μ m, 300 μ m id \times 5 mm) by an autosampler (Surveyor) at a flow of 20 μ L/min for desalting and concentration. The trapped peptides were then back-flushed and separated on a homemade column (length 100 mm) packed with C18 resin (Aqua, 5 μ m, Phenomenex) in 75 mm silica tubing (8 mm id orifice). The mobile phases A and B were composed of 0 and 80% ACN, respectively, each containing 0.5% acetic acid and 0.02% formic acid. The gradient began at 5% B for 15 min; was ramped to 20% B for 3 min, 50% for 47 min, and 95% for 2 min; and held at 95% for 5 min, then 5% for 2 min. The column was equilibrated with 5% B for 6 min before the next run. The eluted peptides were directly electrosprayed into a mass spectrometer, which was controlled by Xcalibur software (Thermo-Electron Corporation, Home Page Version 2.0 SR1). During the gradient elution, three IT MS/MS spectra were acquired *per* data-dependent cycle from a high-resolution (*R* set at 100 000) FT-ICR master spectrum. Ions selected for MS/MS were dynamically excluded for 60 s. Peak lists were generated, and the resulting .raw files were converted to .xml files in Bioworks software (ThermoElectron, ver. 3.3). The.xml files were used for protein identification using MASCOT search engine version 2.0 (Matrix Science) against the IPI human database 20070905 (67 524 sequences; 28 722 560 residues). MASCOT was used with monoisotopic mass selected, a precursor mass tolerance of \pm 1.5 Da, and a fragment mass tolerance of \pm 0.8 Da. Trypsin was selected as the enzyme, with one potential missed cleavage. Oxidized methionine, pyro-glutamate (N-term Q), propionamide cystein and carbamidomethylated cystein was chosen as variable modifications. With regard to acceptance criteria for protein identification, proteins that were identified with more than two peptides among which at least one peptide shows MASCOT individual ion score more than 42 ($p < 0.05$) were considered to be candidates. Criteria that proteins must be a known glycoprotein or otherwise have at least one *N*-glycosylation consensus motif (N-X-S/T, where X is any amino acid except proline) were also considered.

Our experiments constituted the following steps: partial purification of serum proteins, lectin capture and precipitation, tryptic digestion, and protein identification *via* MS and database searches. As diagrammed in Fig. 1A, the overall workflow consisted of two separate strategies: one for optimizing protocols for L-PHA-based lectin capture and another for discovery of CRC biomarkers based on the optimized protocol. Protocol optimization included selection of enrichment or fractionation methods, selection of the form of L-PHA conjugates to be used for capturing β 1,6-GlcNAc-containing *N*-glycoproteins, and determination of the steps for tryptic digestion. Each step that improved the end-result was chosen and combined to generate an “optimized protocol” for biomarker discovery. A primary reason as to why the use of L-PHA is inappropriate for chromatography originates from the unavailability of a nondenaturation-based elution method, as opposed to other widely used lectins such as Con A, *Lens culinaris* agglutinin, and wheat germ agglutinin (WGA), which has necessitated the development of a method to detach L-PHA-bound proteins under denaturing conditions. To this end, L-PHA bound proteins were denatured

in either 6 M urea or SDS-PAGE denaturation buffer. In the solution digestion method, protein preparations denatured in 6 M urea were diluted ten-fold and subjected to tryptic digestion. Meanwhile, protein preparations denatured in denaturation buffer were resolved on an 8% SDS-PAGE gel, after which the entire gel was cut into pieces and in-gel tryptic digestion was performed. We established criteria to determine which digestion method was more effective for protein identification, prioritized by the following criteria: (i) the number of proteins identified from at least two nonoverlapped peptides among which at least one peptide has a significant MASCOT score ($p < 0.05$), (ii) sequence coverage, and (iii) total score values. As is seen in Figs. 1B and C, we identified 28 and 41 proteins satisfying these criteria from the in-gel and solution digestion methods, respectively. In addition, proteins identified through the solution digestion method had higher total scores and sequence coverage compared with those identified through in-gel digestion. Thus, we chose the solution digestion method as described above for subsequent protocol optimization and biomarker discovery.

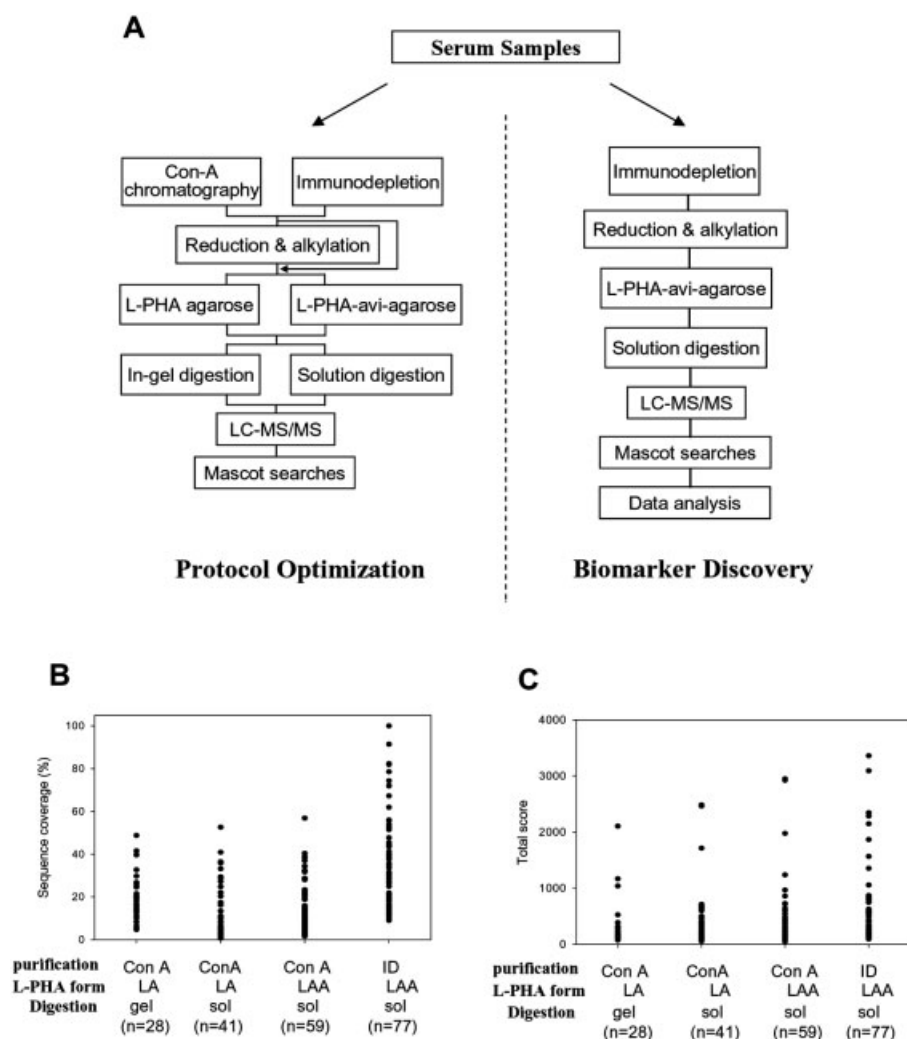


Figure 1. Strategy for protocol optimization of biomarker discovery using serum. (A) Partial purification methods (Con A chromatography or ID), L-PHA bound forms (L-PHA agarose or L-PHA-avidine-agarose), and tryptic digestion methods (in-gel digestion or solution digestion) were varied to monitor the effects on protein identification (left panel). Methods that produced better results were selected and combined for biomarker discovery, as outlined in the right panel. (B, C) Partial purification, L-PHA bound form, and tryptic digestion methods were varied, and the effects on sequence coverage (B) and MASCOT score (C) of the identified proteins from a normal serum sample were investigated. Based on the result, ID as a partial purification method, L-PHA-avidine-agarose, and solution digestion methods (sol) were selected and comprised the optimized protocol for biomarker discovery. A combination of these methods allowed for a ~2.8-fold increase in the number of proteins identified with more than one significant ion score ($p < 0.05$). The results were derived from an experiment showing the best result among three independent experiments.

Another issue related to optimization was deciding which form of L-PHA conjugates – direct conjugation of L-PHA to glycan-based beads such as agarose or sepharose, or indirect conjugation of L-PHA to beads with an intermediate protein – would yield better results. For direct conjugation of L-PHA to beads, reactive group-activated conjugation of the lectin to beads was possible. Instead of performing the conjugation in our laboratory, we purchased commercially available L-PHA-bound agarose beads (L-PHA–agarose) (Sigma) for consistency and reproducibility of results. Indirect conjugation was performed by incubating biotin-labeled L-PHA with avidin-coupled agarose beads (L-PHA–avidine–agarose). Following enrichment of glycoproteins on a Con A-agarose column and preclearing with agarose or avidin–agarose beads, the partially purified serum samples were subjected to lectin precipitation using either L-PHA–agarose or L-PHA–avidine–agarose. The bound proteins were tryptic-digested in solution, and peptide sequencing was performed using an FT-ICR/LTQ mass analyzer. Figures 1B and C show that L-PHA–avidine–agarose was much more effective in capturing target proteins in sera. More proteins were captured by L-PHA–avidine–agarose compared with L-PHA–agarose, most of which were identified by higher total scores and sequence coverage. Steric hindrance and restricted flexibility may be an obstacle for biomolecular interactions, especially for ligand–receptor interactions. Thus, spacer arm-appended matrices are often used to overcome these restrictions in affinity chromatography. Avidin can act as a spacer arm in L-PHA–avidine–agarose complexes and lessen the hindrances that are presumably generated between agarose beads and serum glycoproteins.

Biomarker discovery using plasma or serum is hampered primarily because of the high dynamic range of serological proteins. Albumin accounts for ~45% of total serum proteins, and when combined with Ig, accounts for ~60% of them. In addition, since almost all highly abundant proteins are glycoproteins, depletion of such proteins is demanding, precluding the identification of low-abundance proteins with high accuracy. To the best of our knowledge, the Proteom-Prep 20 LC column (Sigma), a commercially available ID column, removes the highest number of high-abundance proteins. When run on the column, high-abundance proteins were effectively resolved from the remaining serological proteins (Fig. 2A). The pass-through fractions contained mostly low-abundance proteins, from which most of the 20 high-abundance proteins were separated. When 50 µg of proteins were loaded on an SDS-PAGE gel and stained with CBB, albumin was notably visible together with Igs. However, equal amounts of the depleted proteins shows little trace of the high-abundance proteins (Fig. 2B). More importantly, the depleted protein samples were found to be more responsive to L-PHA compared with the undepleted batches (Fig. 2C). Though the amounts of L-PHA-captured proteins are so small that they are invisible in the coomassie stained gel (Fig. 2D), several proteins were differentially displayed between normal and cancer serum in an L-PHA blot analysis

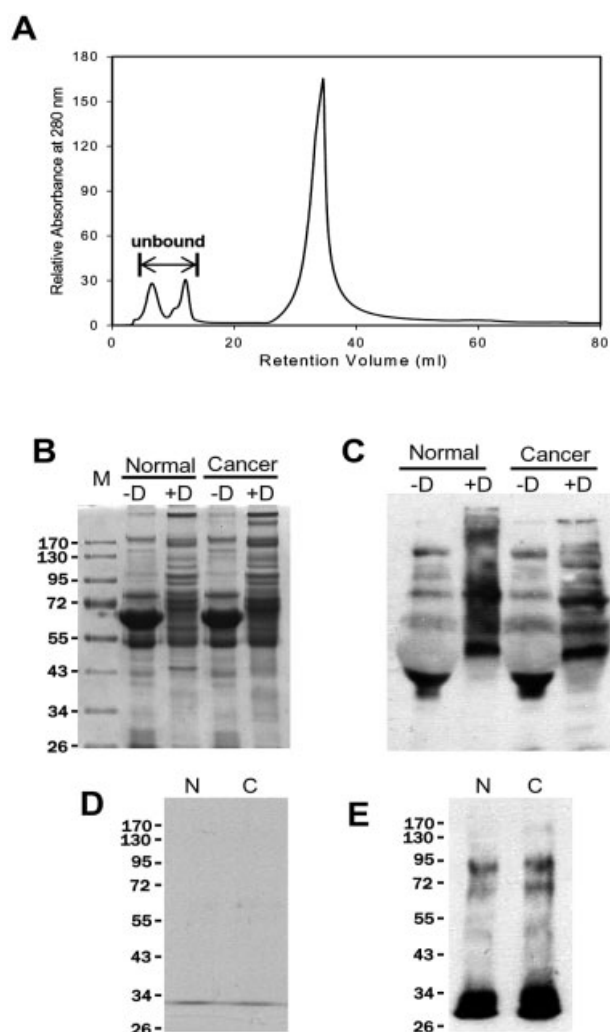


Figure 2. ID of 20 high-abundance proteins from serum. (A) Twenty high-abundance serum proteins were separated from low-abundance proteins (unbound) on a ProteomPrep 20 plasma ID LC column, and the unbound proteins were used for subsequent procedures. (B) Depleted serum samples (+D) were compared with the undepleted ones (-D) on an SDS-PAGE gel following CBB staining. High-abundance proteins such as albumin (~67 kDa) and Ig (~56 kDa) were almost depleted. (C) Equal amounts of serum samples (50 µg proteins) were separated on an SDS-PAGE gel and blotted against L-PHA. After removal of high-abundance proteins, the remaining proteins were more responsive to L-PHA, suggesting that more expansive identification of GnT-V targets is facilitated by ID. (D, E) The immunodepleted protein preparations of normal (N) and cancer (C) sera were captured by L-PHA-avidin-agarose bead complexes, the captured proteins were run on 10% SDS-PAGE gel, stained by CBB (D) and lectin-blotted against L-PHA (E), in which an X-ray film was exposed to chemiluminescence for sufficient time.

(Fig. 2E). In addition, this ID step enabled us to identify more proteins with higher total score values and sequence coverages (Figs. 1B and C). Thus, we implemented ID as a first step in our discovery protocol, which was compatible

with the following step, obviating the need to insert an additional procedure such as buffer exchange that might cause further loss of proteins. Taken together, our results rendered an optimization protocol in which serum samples are immunodepleted and then subjected to denaturation and alkylation steps. These pretreated proteins are bound to L-PHA-avidine-agarose conjugates, and the bound form proteins are denatured in 6 M urea and subjected to tryptic digestion and protein identification.

Clinical serum samples used in our study were grouped into “normal” and “patient” samples. Normal blood samples were taken from six volunteers (mean age, 44 years) who had no previous record of occurrence of any cancer and showed no abnormal signs in medical tests. Patient samples were taken from ten colorectal patients whose demographics and clinical parameters were as follows: mean age of 49 years, regional lymph node metastasis (pN0: 40%, pN1: 50%, pN2: 10%), depth of invasion (pT2: 10%, pT3: 80%, pT4: 10%), and TNM stage (stage I: 10%, stage II: 20%, stage III: 70%). In this study, we adopted a semiquantitative approach: pro-

teins that were captured and detected in patient samples, but not in normal ones, were screened using Microsoft Excel. Among those proteins, proteins identified by only one peptide were eliminated. More than a hundred of proteins satisfied both criteria, from which we narrowed down to 26 candidate biomarkers by screening (i) glycoproteins or proteins that had at least one *N*-glycosylation consensus sequence (N-X-S/T, where X is any amino acid except proline), and (ii) proteins that were identified with a sensitivity of more than 50%. Proteins that passed these filters are listed in Table 1. Because these proteins were not identified in any normal samples, they showed a specificity of 100% in this study.

Oncological processes accompany bilateral changes of biomolecules *i.e.*, qualitative and quantitative. A quantitative change refers to an alteration in expression, especially of proteins, and accordingly, the majority of biomarker discovery studies has relied on the expressional disturbance or fluctuation of proteins throughout a pathological process. Qualitative changes include phosphorylation/dephosphorylation, proteolytic cleavage, glycosylation, and other

Table 1. Lists of biomarker candidates for CRC

Accession no.	Identities	MASCOT score ^{a)}	No. of unique peptides	Sequence coverage (%)	Sensitivity (%)	Specificity (%)
IPI00013835	Isoform long of diacylglycerol kinase zeta	87–196	4–8	7.0–12.2	90	100
IPI00166020	JmjC domain-containing histone demethylation protein 2A	88–135	5–9	7.1–15.2	90	100
IPI00001516	Isoform long of protocadherin α C2 precursor	99–187	2–5	2.3–9.5	80	100
IPI00718805	Brain-rescue-factor-1	86–159	5–10	12.9–35.0	80	100
IPI00002469	Host cell factor 2	96–142	3–6	6.0–13.0	70	100
IPI00021733	Bifunctional heparan sulfate <i>N</i> -deacetylase/ <i>N</i> -sulfotransferase 4	88–141	3–5	5.8–10.1	70	100
IPI00235481	Polyamine modulated factor 1-binding protein 1	96–147	7–11	10.3–18.4	70	100
IPI00293057	Isoform 2 of carboxypeptidase B2 precursor	92–301	2–6	15.8–41.9	70	100
IPI00418544	FLJ00268 protein (fragment)	86–143	4–7	9.8–20.3	70	100
IPI00739954	Similar to α 3 type VI collagen isoform 1 precursor	86–150	7–14	23.6–48.7	70	100
IPI00786924	Similar to B0416.5a	85–134	7–10	24.3–30.2	70	100
IPI00787145	OTU domain containing 1	98–191	6–14	17.2–46.4	70	100
IPI00009793	Complement C1r-like protein	90–263	2–4	9.4–16.2	60	100
IPI00019223	Isoform 1 of A-kinase anchor protein 9	101–158	8–15	4.4–7.7	60	100
IPI00164623	187 kDa protein	154–875	5–8	6.2–10.2	60	100
IPI00169307	Rho-GTPase activating protein 10	86–207	4–8	3.4–7.8	60	100
IPI00218823	Isoform 1 of WW domain-binding protein 7	136–232	20–27	19.8–30.1	60	100
IPI00291638	183 kDa protein	86–117	5–9	5.7–10.5	60	100
IPI00452247	Isoform 2 of kinesin-like motor protein C20orf23	90–138	5–8	6.0–11.2	60	100
IPI00477931	361 kDa protein	91–160	4–10	2.0–6.5	60	100
IPI00553067	Coiled-coil domain containing 132 isoform a	88–152	4–7	7.7–14.4	60	100
IPI00746049	Similar to breast cancer antigen NY-BR-1.1	105–152	7–13	10.5–34.7	60	100
IPI00854651	Isoform 1 of formin-1	101–140	4–7	3.5–10.7	60	100
IPI00021274	Ephrin type-A receptor 8 precursor	104–165	6–9	10.9–15.8	50	100
IPI00163493	Isoform 1 of aminopeptidase O	86–106	5–7	17.8–24.3	50	100
IPI00299059	Isoform 2 of neural cell adhesion molecule L1-like protein precursor	135–390	2–7	3.0–13.8	50	100

a) Score is $-10 \times \log(p)$, where p is the probability that the observed match is a random event.

PTMs. In particular, glycosylation changes have been reported to be associated with malignant transformation, invasion, and metastasis [4]. In this regard, a concomitant consideration of qualitative and quantitative changes in biomarkers can provide us with more useful information for cancer prediction and diagnosis. For example, α -fetoprotein (AFP) has been used as the sole biomarker for hepatocellular carcinoma, but its use is limited by its low sensitivity (~50%). However, when α 1,6-fucosylation on AFP is considered together with the titer elevation, the sensitivity increases to 69% and the specificity was 96% in benign liver diseases [11]. Use of L-PHA as a capture lectin enables us to measure both qualitative and quantitative changes in proteins, because an increase in capture by L-PHA reflects a net increase in the β 1,6-GlcNAc glycan moiety without any significant change in expression level. Such a signal, however, can also result simply from the upregulation of otherwise undetectable glycoproteins with a cognate β 1,6-GlcNAc linkage or from simultaneous increases in β 1,6-GlcNAc glycan moiety and protein levels.

Most glycoproteins either are retained on the membrane fraction (*e.g.*, plasma membrane and ER membrane) or flow out of cells by secretion and shedding. For this reason, serum is a depository for glycoproteins originating from multiple tissues, and many useful biomarkers such as PSA and AFP are in fact glycoproteins. This utility of glycoproteins as potential biomarkers has led to the use of various lectins for up-front capture of glycoproteins, and as such, lectin capture has been recognized as a robust step for identifying serological glycoprotein biomarkers when combined with MS [12]. *Aleuria aurantia* lectin (AAL), *Anguilla anguilla* agglutinin (AAA), and *Sambucus nigra* (SNA1) belong to a lectin group that recognizes a specific glycan moiety. Con A and WGA are two examples of lectins with broad coverage of glycoproteins. The aforementioned lectins may be suitable for unbiased biomarker discovery, but an additional fractionation step would be needed for *de novo* biomarker discovery using serum because evidence indicating that the glycan branches recognized by the lectins are implicated in the oncopathology is scanty. Moreover, Con A catches nearly all *N*-glycosylated proteins, and if the lectin is used, much more lectin is needed to capture glycoproteins in a specified volume of serum, and high amounts of lectin surely interfere with identification of proteins of interest during MS analysis. The restrictions of these lectins when used in glycoprotein enrichment propelled us to conduct “pathology-implicated targeted capture” using L-PHA, which characteristically binds with high affinity to β 1,6-GlcNAc on *N*-glycoproteins generated by GnT-V activity [13].

Table 1 lists 26 candidate biomarkers for CRC, most of which have not been reported as biomarkers except for NY-BR-1.1. NY-BR-1.1 is a homolog of NY-BR-1, which was previously identified by the SEREX method [14] and is sporadically expressed in 32% of breast tumors [15]. To test if the candidate proteins are specific as biomarkers for CRC patients, we performed biomarker discovery using sera of

liver cancer patients following the same protocol developed in this study. Liver cancer is chosen because it is a representative cancer that shows high expression of GnT-V [16, 17]. As a result, we could obtain a list of candidate biomarkers for liver cancer comprising 24 proteins with 100% specificity and more than 50% sensitivity (data not shown). Interestingly, most of biomarker candidates for liver cancer did not overlap with those for CRC except for polyamine modulated factor 1-binding protein 1 (IPI00235481) and isoform 1 of formin-1 (IPI00854651), which probably reflect that each of the candidate proteins is specific biomarkers for the corresponding cancer type. As biomarker discovery innately generates a panel of proteins with a high false discovery rate and low credentialing [18], our list of proteins should be tentatively considered “candidate biomarkers” for CRC. These proteins await further validation steps in order to distinguish the true positives. Additionally, our optimized protocol may be directly applicable to the discovery of candidate biomarkers for other, nonCRCs.

This work was supported by the “21st Century Frontier R&D Program for Functional Analysis of Human Genome,” the “Leading Foreign Research Institute Recruitment Program,” the “Complex Carbohydrate Research Program,” and the “STRM Program” from the Korea Ministry of Science and Technology. Especially, we express deep appreciation to Daejeon metropolitan city for its support.

The authors have declared no conflict of interest.

References

- [1] Parkin, D. M., Pisani, P., Ferlay, J., Global cancer statistics. *CA Cancer J. Clin.* 1999, 49, 33–64.
- [2] Petersen, G. M., Brensinger, J. D., Johnson, K. A., Giardiello, F. M., Genetic testing and counseling for hereditary forms of colorectal cancer. *Cancer* 1999, 86, 2540–2550.
- [3] Powell, S. M., Petersen, G. M., Krush, A. J., Booker, S., Jen, J. *et al.*, Molecular diagnosis of familial adenomatous polyposis. *N. Engl. J. Med.* 1993, 329, 1982–1987.
- [4] Orntoft, T. F., Vestergaard, E. M., Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* 1999, 20, 362–371.
- [5] Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M. L., Kerbel, R. S., Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* 1987, 236, 582–585.
- [6] Ihara, S., Miyoshi, E., Ko, J. H., Murata, K. *et al.*, Prometastatic effect of N-acetylglucosaminyltransferase V is due to modification and stabilization of active matriptase by adding beta 1–6 GlcNAc branching. *J. Biol. Chem.* 2002, 277, 16960–16967.
- [7] Guo, H.-B., Lee, I., Kamar, M., Akiyama, S. K., Pierce, M., Aberrant N-glycosylation of β 1 integrin causes reduced α 5 β 1 integrin clustering and stimulates cell migration. *Cancer Res.* 2002, 62, 6837–6845.

- [8] Guo, H.-B., Lee, I., Kama, M., Pierce, M., N-acetylglucosaminyltransferase V expression levels regulate cadherin-associated homotypic cell-cell adhesion and intracellular signaling pathways. *J. Biol. Chem.* 2003, *278*, 52412–52424.
- [9] Kim, Y.-S., Hwang, S. Y., Kang, H.-Y., Sohn, H. *et al.*, Functional proteomics study reveals that N-acetylglucosaminyltransferase V reinforces the invasive/metastatic potential of colon cancer through aberrant glycosylation on tissue inhibitor of metalloproteinase-1. *Mol. Cell. Proteomics* 2008, *7*, 1–14.
- [10] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 1996, *68*, 850–858.
- [11] Aoyagi, Y., Suzuki, Y., Isemura, M., Nomoto, M. *et al.*, The fucosylation index of alpha-fetoprotein and its usefulness in the early diagnosis of hepatocellular carcinoma. *Cancer* 1988, *61*, 769–774.
- [12] Drake, R. R., Schwegler, E. E., Mali, G., Diaz, J., Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* 2006, *5*, 1957–1967.
- [13] Cummings, R. D., Kornfeld, S., Characterization of the structural determinants required for the high affinity interaction of asparagine-linked oligosaccharides with immobilized Phaseolus vulgaris leucoagglutinating and erythroagglutinating lectins. *J. Biol. Chem.* 1982, *257*, 11230–11234.
- [14] Jäger, D., Stockert, E., Güre, A. O., Scanlan, M. J. *et al.*, Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res.* 2001, *61*, 2055–2061.
- [15] Jäger, D., Karbach, J., Pauligk, C., Seil, I. *et al.*, Humoral and cellular immune responses against the breast cancer antigen NY-BR-1: Definition of two HLA-A2 restricted peptide epitopes. *Cancer Immun.* 2005, *5*, 11.
- [16] Miyoshi, E., Nishikawa, A., Ihara, Y., Gu, J. *et al.*, N-acetylglucosaminyltransferase III and V messenger RNA levels in LEC rats during hepatocarcinogenesis. *Cancer Res.* 1993, *53*, 3899–3902.
- [17] Ito, Y., Miyoshi, E., Sakon, M., Takeda, T. *et al.*, Elevated expression of UDP-N-acetylglucosamine: Alphanmannoside beta1,6 N-acetylglucosaminyltransferase is an early event in hepatocarcinogenesis. *Int. J. Cancer* 2001, *91*, 631–637.
- [18] Rafai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* 2006, *24*, 971–983.