

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 June 2009 (18.06.2009)

PCT

(10) International Publication Number
WO 2009/075883 A2

(51) International Patent Classification:
G01N 33/574 (2006.01)

(21) International Application Number:
PCT/US2008/013658

(22) International Filing Date:
12 December 2008 (12.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/007,442 12 December 2007 (12.12.2007) US

(71) Applicant (for all designated States except US): **UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC.** [US/US]; Boyd Graduate Studies Research Center, Athens, GA 30602-7411 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PIERCE, Michael, J.** [US/US]; 225 Garland Hill Drive, Athens, GA 30602 (US). **ABBOTT, Karen, L.** [US/US]; 1140 Derby Lane, Watkinsville, GA 30677 (US).

(74) Agent: **SANDBERG, Victoria, A.**; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581336, Minneapolis, MN 55458-1336 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

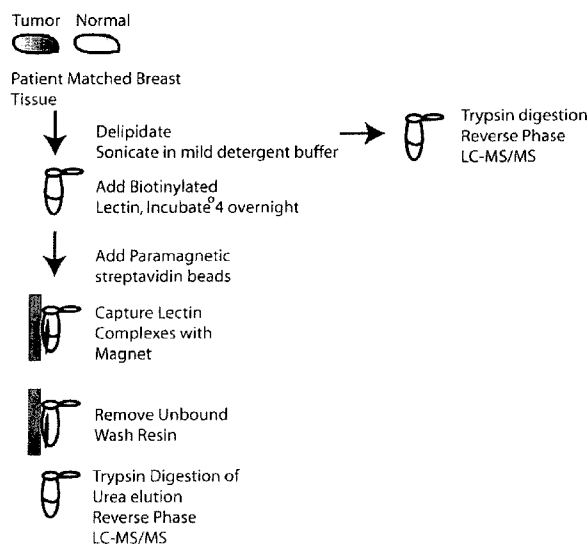
Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: GLYCOPROTEIN CANCER BIOMARKER

Fig. 3



(57) Abstract: The invention provides method for identifying glycoproteins having a cancer-specific glycoform. Cancer-specific glycoforms are useful in diagnostics and therapeutics.

WO 2009/075883 A2



-
- *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

5

GLYCOPROTEIN CANCER BIOMARKER

CONTINUING APPLICATION DATA

This application claims the benefit of U.S. Provisional Application Serial No.
10 61/007,442, filed December 12, 2007, which is incorporated by reference herein.

BACKGROUND

15 Breast carcinoma is the second leading cause of cancer deaths among women in the U.S. (Jemal et al., 2005. *CA Cancer J Clin* 55:10-30). Early detection and diagnosis of breast cancer significantly improves 5-year survival rates (Ries et al., 2005. SEER Cancer statistics review. National Cancer Institute, Bethesda, MD). Currently the only approved screening method for the detection for breast cancer is mammography.

20 In the past few years, several large-scale proteomic studies have begun to characterize the proteome of breast cancer (Pucci-Minafra et al., 2006. *Proteomics* 6:2609-25; Celis et al., 2004. *Mol Cell Proteomics* 3:327-44; Hudelist et al., 2006. *Proteomics* 6:1989-2002). This high-throughput strategy leads to complex data sets that, while rich in information, are often not very useful in predicting proteins that may be
25 sensitive and specific biomarkers for the disease.

Epithelial ovarian cancer is the deadliest reproductive tract malignancy of women in Western countries (Ozols et al., 2004. *Cancer Cell* 5:19-24). Ovarian cancer survival rates at 5 years are only 30% for women diagnosed with distant metastases; however, the percentage survival climbs to 90% for women diagnosed with disease confined to the
30 ovary (Hayat et al., 2007. *Oncologist* 12:20-37). Unfortunately, fewer than 25% of women are diagnosed when the disease is confined, due primarily to the lack of screening tests capable of detecting ovarian cancers early.

Reliable biomarkers are not available for the majority of cancers and precancerous conditions. Specifically, there is a lack of secreted biomarkers that can be detected
35 through non-invasive assays such as blood tests. Without convenient and easily accessible screening tests for cancers and precancerous conditions, diagnostic delays will

continue to plague the health care system and thwart efforts to detect and treat malignancies in their earliest stages.

5

SUMMARY OF THE INVENTION

The present invention is directed to cancer biomarkers, particularly glycoprotein biomarkers. A preferred cancer biomarker according to the invention is a cancer-specific glycoform of a glycoprotein. The present invention additionally provides methods for
10 identifying glycoproteins possessing a cancer-specific glycoform, as well as diagnostic and therapeutic methods and compositions related to glycoprotein cancer biomarkers. Cancer biomarkers of the invention can be specific for any cancer, without limitation. Preferred cancer biomarkers are specific for breast cancer, ovarian cancer, colorectal cancer, pancreatic cancer and liver cancer.

15 In one aspect, the invention provides a diagnostic method for evaluating the presence, absence, nature or extent of cancer or a precancerous condition in a subject. The subject can be any mammalian subject, and is preferably a human or a domestic animal. In one embodiment, the diagnostic method involves detecting the presence of a cancer-specific glycoform of a glycoprotein in a biological sample obtained from the
20 subject, wherein the cancer-specific glycoform comprises glycan that is indicative of the presence of cancer or a precancerous condition. The biological sample can include, without limitation a biological fluid, such as blood, serum or plasma, or it can be a tissue or organ sample.

Any convenient method can be used to detect the cancer-specific glycoform. In
25 one embodiment, the cancer-specific glycoform can be detected by contacting the biological sample with a glycan-binding molecule specific for the glycan, under conditions that permit binding of the cancer-specific glycoform of the glycoprotein to the glycan-binding molecule. Exemplary and preferred glycan-binding molecules include a lectin, a glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, and a
30 glycospecific small molecule. In other embodiments, other detection methods can be used, for example mass analysis methods such as mass spectrometry.

In an embodiment of the diagnostic method wherein cancer or a precancerous condition of the breast is evaluated, exemplary glycoprotein breast cancer biomarkers include glycoproteins set forth in Table 4, including periostin and osteoglycin. Examples
35 of glycans that can be detected on a breast cancer-specific glycoprotein glycoform include

a GlcNAc β (1,6) Man branched *N*-linked glycan and a branched *N*-linked glycan extended with N-acetyllactosamine. A preferred glycan-binding molecule for use in evaluating breast cancer is the lectin leucoagglutinating phytohemagglutinin (L-PHA).

In an embodiment of the diagnostic method wherein cancer or a precancerous condition of the ovary is evaluated, exemplary glycoprotein ovarian cancer markers include glycoproteins set forth in Table 8. Examples of glycans that can be detected on an ovarian cancer-specific glycoprotein glycoform include a glycan containing α (1,6)-fucose linked to core *N*-acetylglucosamine (core fucosylation), and a GlcNAc β (1,4) Man bisected *N*-linked glycan. Preferred glycan-binding molecules for use in evaluating ovarian cancer include erythroagglutinating phytohemagglutinin (E-PHA), Aleuria aurantia lectin (AAL) and Datura stramonium lectin (DSL).

In another aspect, the invention provides a method for identifying a biomarker associated with cancer. In one embodiment, a biological test sample from a subject having cancer or a precancerous condition is contacted with a glycan-binding molecule specific for a glycan, under conditions that permit binding of the glycan-binding molecule to the glycan, when present, yielding a bound glycoprotein/glycan-binding molecule complex. The biological sample is preferably a tissue or organ sample; for example obtained from a cancerous tumor.

A reference or control sample (a noncancerous biological sample) is preferably analyzed and compared with the test sample from the subject with cancer or a precancerous condition. The noncancerous biological sample can be from the same subject, a different subject, or it can be a pooled sample from a number of disease-free subjects. The noncancerous biological sample can be contacted with the glycan-binding molecule specific for the glycan, under conditions that permit binding of the glycan-binding molecule to the glycan, when present, to yield a bound glycoprotein/glycan-binding molecule complex. The presence of a cancer-specific glycoform can be evaluated by determining whether the amount of bound glycoprotein from the tumor sample is greater than the amount of bound glycoprotein from the noncancerous sample, wherein a greater amount in the tumor sample indicates the presence of a cancer-specific glycoform of the glycoprotein. Preferred glycan-binding molecules include those that are preferred for use in the diagnostic method of the invention, but are not limited thereto. Preferred glycans that can be detected with a glycan-binding molecule according to the identification method of the invention a GlcNAc β (1,6) Man branched *N*-linked glycan, a GlcNAc β (1,4) Man bisected *N*-linked glycan, a glycan containing α (1,6) fucose linked to

a core *N*-acetylglucosamine, and a branched *N*-linked glycan extended with *N*-acetyllactosamine.

In another aspect, the invention includes glycoprotein cancer biomarkers identified using methods as described herein. Exemplary glycoprotein cancer biomarkers are set forth in Tables 4 and 6, and include a cancer-specific glycoform of periostin, preferably including a GlcNAc β (1,6) Man branched *N*-linked glycan component, and a cancer-specific glycoform of osteoglycin, preferably including GlcNAc β (1,6) Man branched *N*-linked glycan component. Cancer-specific glycoforms of glycoproteins included in the present invention include but are not limited to cancer-specific glycoforms possessing one or more of the following glycan components or structural features: a GlcNAc β (1,6) Man branched *N*-linked glycan component and a branched *N*-linked glycan extended with *N*-acetyllactosamine, a GlcNAc β (1,4) Man bisected *N*-linked glycan component, and an α (1,6) fucose linked to a core *N*-acetylglucosamine.

The diagnostic methods of the invention and the biomarker identification method of the invention are both amenable to multiplexing. For example, a multiplexed diagnostic method for evaluating the presence, absence, nature or extent of cancer or a precancerous condition in a subject can include providing a biological sample obtained from the subject, wherein the biological sample includes a plurality of glycoproteins; and, for each of the plurality of glycoproteins, determining the presence, absence or amount of a cancer-specific glycoform of the glycoprotein in the biological sample, wherein the presence, absence or amount of the cancer-specific glycoform is indicative of cancer or a precancerous condition. In a multiplexed breast cancer diagnostic method, preferably the plurality of glycoproteins evaluated includes at least two proteins independently selected from the proteins set forth in Table 4. More preferably, the multiplexed breast cancer detection method detects cancer-specific glycoforms of periostin and osteoglycin. In a multiplexed ovarian cancer diagnostic method, preferably the plurality of glycoproteins evaluated includes at least two proteins independently selected from the proteins set forth in Table 8.

The words "preferred" and "preferably" refer to embodiments of the invention that may afford certain benefits, under certain circumstances. However, other embodiments may also be preferred, under the same or other circumstances. Furthermore, the recitation

of one or more preferred embodiments does not imply that other embodiments are not useful, and is not intended to exclude other embodiments from the scope of the invention.

The terms "comprises" and variations thereof do not have a limiting meaning where these terms appear in the description and claims.

5 Unless otherwise specified, "a," "an," "the," and "at least one" are used interchangeably and mean one or more than one.

Also herein, the recitations of numerical ranges by endpoints include all numbers subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, 5, etc.).

10 For any method disclosed herein that includes discrete steps, the steps may be conducted in any feasible order. And, as appropriate, any combination of two or more steps may be conducted simultaneously.

The above summary of the present invention is not intended to describe each disclosed embodiment or every implementation of the present invention. The description that follows more particularly exemplifies illustrative embodiments. In several places
15 throughout the application, guidance is provided through lists of examples, which examples can be used in various combinations. In each instance, the recited list serves only as a representative group and should not be interpreted as an exclusive list.

20

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a lectin blot demonstrating the reactivity of L-PHA toward proteins extracted from wild-type and GnT-V(-) MDA-MB231 invasive breast carcinoma
25 cells.

Figure 2 shows a tetra-antennary *N*-linked oligosaccharide showing the GnT-V $\beta(1,6)$ GlcNAc addition that leads to the formation of polylactosamine structures. The L-PHA recognition site is circled.

Figure 3 shows a schematic flow diagram for the L-PHA enrichment protocol (B).

30 Figure 4 shows enrichment of tetra-antennary glycans extended with *N*-acetylactosamine in tumor and adjacent normal tissue from case 2417. The four indicated *N*-linked glycans (1 – 4) were detected by NSI-MS/MS. In the profiles shown at the right of each glycan, the MS/MS spectra associated with the TIM scan for the indicated tissue were filtered to present the detected signal intensity of a signature tetrasaccharide
35 fragment (Hex-HexNAc-Hex-HexNAc). The presence of this fragment indicates the

detection of a glycan extended with at least two *N*-acetylglucosamine repeats at a scan time which predicts the *m/z* ratio for the parent ion. For reference, the scan time for specific *m/z* values are indicated by arrows in each filtered profile. The shading and shapes for the glycan structures reflect standard nomenclature adopted by the Consortium for Functional Glycomics (CFG; N-Acetylglucosamine, GlcNAc, square; galactose, Gal, open circle, mannose, Man, filled circle; fucose, Fuc, triangle, N-Acetylneuraminic acid, NeuAc, diamond).

Figure 5 shows the functional annotation and distribution of the L-PHA enriched proteins identified from breast carcinoma. (A) Biological function of proteins listed in Table 2 as annotated by DAVID 2007 (B) Cellular compartment for L-PHA enriched proteins assigned based on GO consortium (C) Cellular compartment of proteins identified from normal breast tissue by total MS/MS analysis assigned by the Gene Ontology (GO) consortium. (D) Cellular compartment of proteins identified from tumor breast tissue assigned by GO consortium.

Figure 6 shows a Venn diagram showing the number of L-PHA enriched proteins identified in common for each case.

Figure 7 shows an analysis of periostin and haptoglobin-related protein by Western blot. (A) Number of peptides identified for periostin before L-PHA fractionation (total) and after lectin fractionation (L-PHA). (B) Precipitation of periostin using an anti-periostin antibody followed by detection using biotinylated L-PHA and streptavidin HRP (panel 1). Total levels of periostin precipitated are confirmed by detection of the blot using anti-periostin antibody (panel 2). Reverse precipitation with L-PHA first followed by detection using an anti-periostin antibody (panel 3). (C) Densitometry quantification of the relative increase in L-PHA reactive periostin normalized for total periostin. (D) Number of peptides identified for haptoglobin-related protein (HPR) precursor by MS/MS before (total) and after L-PHA fractionation (L-PHA). (E) L-PHA precipitation followed by detection using an anti-haptoglobin antibody shows increased reactivity for the beta chains of tumor HPR for cases 2417 and 2207 and in all cases a migratory shift to a higher molecular weight. (F) Relative levels of L-PHA reactive haptoglobin were determined following densitometric analysis normalizing to the total levels of haptoglobin on 10 % input blots (data not shown).

Figure 8 shows transcript analysis of enzymes acting in the *N*-linked pathway for normal ovary. (A) *N*-linked glycosylation pathway with enzymes included in analysis numbered as follows: 1, MGAT1; 2, MAN2A1 (Man II) or MAN2A2 (Man IIx); 3,

---FUT8; 4, MGAT2; 5, MGAT3; 6, MGAT4a or MGAT4b; 7, MGAT5. (B) Relative transcript levels for normal human ovary tissue, average Ct for two pooled cases. Error bars represent the SD from the mean for triplicate Ct values. (C) Relative transcript levels for normal mouse ovary, average Ct for six pooled normal ovaries. Error bars represent the SD from the mean for triplicate Ct values.

Figure 9 shows comparative analysis of normal and endometrioid ovarian carcinoma. (A) Relative transcript abundance for mouse normal and ovarian tumors plotted on a log scale. Error bars represent the SD from the mean for triplicate Ct values. (B) Relative transcript abundance for human normal and ovarian tumors plotted on a log scale. Error bars represent the SD from the mean for triplicate Ct values.

Figure 10 shows oligosaccharides determinants for lectin binding affinity. Examples of structures with important binding determinants from each lectin circle.

Figure 11 shows lectin blot analysis of glycoproteins from mouse and human endometrioid ovarian carcinoma. (A) Glycoproteins extracted from mouse endometrioid ovarian tumors (lanes 2, 4, 6, and 8) or normal mouse ovary (lanes 1, 3, 5, and 7) that were nonadherent to Con A were separated on 4–12% Bis-Tris gels before transfer to PVDF membrane and detection using biotinylated lectins and streptavidin–HRP. Panel below shows the densitometry analysis of bands from normal (NL) or ovarian tumor (OT) in the 49–250 kDa range from the blots shown above with normal set at 1.0 for comparison. Fold increase was adjusted for lectin pull-down inputs based on the levels of ERK2 on a 10% input blot (data not shown). (B) Glycoproteins nonadherent to Con A from human endometrioid ovarian cancer cases (711, 741, and 471) and normal human ovary (NL) were separated on 4–12% Bis-Tris gels before lectin blot detection as described. The panel below represents the densitometry results for glycoproteins 49–250 kDa relative to normal set at 1.0. Increases relative to normal were adjusted for input using ERK2 analysis form 10% input blots (data not shown).

Figure 12 shows a Western blot, using antibodies against periostin (POSTN) and osteoglycin (OGN), probed with L-PHA enriched serum from patients with breast cancer (BC) and healthy patients (NL).

Figure 13 shows Western blots for three different proteins immunoprecipiated from ovarian tumor tissue (TU) and benign ovarian tissue (NL), and probed with lectins DSL, AAL and E-PHA, (A) lysosomal-associated membrane glycoprotein 1 (LAMP-1), (B) periostin (POSTN), and (C) lectin galactosidase soluble binding protein 3 (GALS3BP).

Figure 14 shows Western blots, using antibodies against (A) periostin and a control, (B) α -1 acid glycoprotein, with E-PHA enriched serum from patients with ovarian cancer (OT) compared with serum from women with benign uterine conditions (OB).

Figure 15 shows evidence of increased fucosylated glycans on lysosomal-associated glycoprotein 1 (Lamp-1) present in serum from patients with ovarian cancer (OT) compared with serum from women with benign uterine conditions (OB).

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention is based upon the discovery that there exist cancer-specific glycoforms of various glycoproteins, and that these glycoforms can be identified, detected, and distinguished through the use of glycan-binding molecules such as lectins. Cancer-specific glycoforms serve as useful biological markers, i.e., biomarkers, of cancer or precancerous conditions. The cancer-specific biomarkers of the invention can be used to detect or monitor the progression of cancer or precancerous conditions, as well as to distinguish cancerous or precancerous conditions from each other and from benign (non-cancerous) disease states. The cancer-specific biomarker may be specific for a disease stage, hormone receptor status, lymph node status and her2/new status. The present invention provides both individual biomarkers as well as a panel of biomarkers useful for cancer screening.

Surprisingly, the invention allows detection of a cancer biomarker protein that may be present in equal or indistinguishable protein levels in diseased and normal samples. Although the total protein level or abundance may be equal as between the two samples, the invention facilitates discrimination between a protein glycoform that is present in the diseased and a different protein glycoform present in the normal sample.

Advantageously, cancer-specific biomarkers of the invention include glycoproteins that are present in blood or blood components, such as serum and/or plasma, or that have been secreted into other biological fluids. Biomarkers that are present in bodily fluids are especially well-suited for high throughput screening and early detection efforts. Blood and other bodily fluids present a readily accessible format for early detection. The identification of fluid biomarkers and development of a non-invasive screening test for cancers and precancerous conditions represents a significant medical advance.

Exemplary cancer-specific biomarkers of the invention include periostin and osteoglycin (also known as mimecan), each of which possesses at least one cancer-specific glycoform.

The invention also provides a method for identifying cancer-specific biomarkers.

5 The identification method of the invention represents a targeted glycoproteomic approach to biomarker identification. The method makes use of a glycan-binding molecule, such as a glycospecific antibody, an a glycospecific aptamer, or a lectin, to identify a glycoform of a glycoprotein that is enriched in a biological sample containing cancer cells compared to a normal biological sample, and is thus cancer-specific. In some instances, the
10 identification method may find a cancer-specific glycoprotein that is not found at all in normal samples. In other instances, the identification method may find a cancer-specific protein that is not a glycoprotein, but associates with a glycoprotein that binds to the glycan-binding molecule and is as a result affinity-enriched. In other instances, the identification method may find a cancer-specific glycoprotein, which is not glycosylated
15 in its normal form but becomes glycosylated during oncogenic transformation. Typically, however, the method identifies a cancer-specific glycoform of a glycoprotein, with different glycoforms of the glycoprotein being found in healthy and cancerous tissues or fluids.

Lectins and other glycan-binding molecules can be used in the diagnostic methods
20 of the invention to detect the presence of cancer-specific biomarker glycoforms, but it should be understood that cancer-specific glycoforms of glycoproteins can be detected using other detection methods as well. Furthermore, in addition to having utility in diagnostic methods, the cancer-specific biomarkers of the invention may serve as or may further suggest additional therapeutic targets.

25

Cancer-specific biomarker

A biomarker is a molecular, biological, or physical characteristic that can be measured or otherwise evaluated as an indicator of a normal biologic process, disease state, or response to a therapeutic intervention. The biomarker of the invention is a
30 “cancer-specific” biomarker, i.e., it is indicative of cancer or a precancerous condition. Biomarkers of the invention include, but are not limited to, an RNA transcript, a protein, or a modified protein such as a glycoprotein. Biomarkers are detectable and/or measurable by any of a variety of methods such as biochemical and/or molecular assays.

The cancer-specific biomarker may be indicative of cancer or a precancerous condition by its presence, absence, increase in amount, decrease in amount, or differential glycosylation. Advances in proteomic methods now enable one to quantify a biomarker such that increases or decreases in the abundance of the biomarker, in addition to the complete presence or absence of the biomarker, may be indicative of the presence of cancer or a precancerous condition. Amounts of biomarker can be determined in absolute or relative terms. Accordingly, a glycoprotein biomarker may indicate the presence of cancer or a precancerous condition simply by its presence, absence or amount compared to a noncancerous sample or a predetermined level; however, the cancer-specific glycoprotein biomarker preferably takes the form of a cancer-specific glycoform of the glycoprotein. In another embodiment, the cancer-specific biomarker is an enzyme that catalyzes the particular linkage characterizing, and preferably specific to, the cancer-specific glycoform.

The cancer-specific biomarker of the invention is preferably a protein, more preferably a post-translationally modified protein, and even more preferably, a glycoprotein. The term "protein," as used herein, refers broadly to a polymer of two or more amino acids joined together by peptide bonds. The term "protein" does not imply a particular polymer length and thus includes peptides, oligopeptides, and polypeptides. The term protein also includes molecules which contain more than one peptide joined by a disulfide bond, or complexes of peptides that are joined together, covalently or noncovalently, as multimers (e.g., dimers, tetramers). The term "glycoprotein" includes any molecule that contains both a protein component and a carbohydrate component. The carbohydrate component is commonly referred to as a "glycan." As used herein, the term glycoprotein is inclusive of a glycopeptide, a glycopolypeptide and a proteoglycan. A glycan may contain one monosaccharide, or it may contain two or more monosaccharides linked by glycosidic bonds. A glycan can include nonrepeating or repeating monosaccharides, or both.

As used herein, the term "glycan" is interchangeable with the term saccharide, which includes a monosaccharide, a disaccharide or a trisaccharide; it can include an oligosaccharide or a polysaccharide. An oligosaccharide is an oligomeric saccharide that contains two or more saccharides. The structure of an oligosaccharide is typically characterized by particular identity, order, linkage positions (including branch points), and linkage stereochemistry (α , β) of the monomers, and as a result has a defined molecular weight and composition. An oligosaccharide typically contains about 2 to

about 20 or more saccharide monomers. In a polysaccharide, the identity, order, linkage positions (including branch points) and/or linkage stereochemistry can vary from molecule to molecule. Polysaccharides typically contain a larger number of monomeric components than oligosaccharides and thus have higher molecular weights. The term
5 "glycan" as used herein is inclusive of both oligosaccharides and polysaccharides, and includes both branched and unbranched polymers as defined herein.

The glycan component of a glycoprotein can be *N*-linked or *O*-linked. An *N*-glycan is attached to a nitrogen atom, for example, at the side chain nitrogen atom of an asparagine amino acid within the peptide. An *O*-linked glycan is attached to an oxygen
10 atom, for example at the side chain hydroxyl oxygen of a hydroxylysine, hydroxyproline, serine, or threonine amino acid within the peptide.

"Glycosylation" refers to the covalent attachment of at least one saccharide moiety to a molecule. Glycosidic linkages include *O*-glycosidic linkages, *N*-glycosidic linkages, *S*-glycosidic linkages and *C*-glycosidic linkages. An *O*-glycosidic linkage is
15 formed between the anomeric carbon (C1) of a saccharide and an oxygen atom of another molecule (such as another saccharide or a polypeptide), while an *N*-glycosidic linkage is formed between the anomeric carbon (C1) of a saccharide and a nitrogen atom of another molecule. Likewise, *S*-glycosidic linkages and *C*-glycosidic linkages involve a sulphur and carbon atom from another molecule, respectively. In addition, glycosidic linkages
20 are classified according to the ring position of the carbon atoms participating in the bond. For example, a 1,4 glycosidic linkage is formed between the first carbon (C1) on a first saccharide and the fourth carbon (C4) on a second saccharide while a 1,6 glycosidic linkage is formed between the first carbon (C1) on a first saccharide and the sixth carbon (C6) on a second saccharide. Glycosidic linkages are further classified as α -glycosidic or
25 β -glycosidic according to whether the substituent groups on the carbons flanking the oxygen in the saccharide are pointing in the same or opposite directions. The term "glycosylation" as used herein should be broadly construed so as to encompass the covalent linkage of any other carbohydrate moieties such as fucose and sialic acid, and as such includes fucosylation or sialylation. Most *N*-linked glycans share a common
30 structure, referred to as a core, which typically contains three mannose, and two *N*-acetylglucosamine residues. The core may contain modifications such as sulfation or phosphorylation; the core may be intact or it may be truncated. Terminal modifications and core modifications of a glycan can include glycosylations. Core glycosylation refers

to the addition of glycosyl moieties to a core N-acetylglucosamine. Core fucosylation refers to the addition of a fucose residue to the core N- acetylglucosamine.

A glycan can be branched or unbranched. A complex glycan is a glycan that contains at least one branch point. In a complex or branched glycan, the monosaccharide at the branch point is covalently linked to two other saccharides at carbons other than C1. For example, a branch point monosaccharide may be linked to other monosaccharides at C4 and C6, in addition to being linked to another monosaccharide or to an amino acid at C1. A complex glycan may be, without limitation, biantennary, triantennary, or tetraantennary. Additionally or alternatively, a complex glycan may be bisected (see, e.g., the structure catalyzed by MGAT3 in reaction 5 in Fig. 8).

Cancer-specific glycoforms

Glycosylation is a dynamic, post-translational modification that can be altered during the development and progression of a cancer or a precancerous condition. As a result, the same glycoprotein ("same" in the sense that it contains the same or essentially the same amino acid sequence; for example protein isoforms are considered to be the same protein) may be expressed both before and after oncogenic transformation, but the glycosylation of the glycoprotein before and after oncogenic transformation may be different. Glycoproteins having the same or essentially the same protein sequence, but exhibiting a difference in glycosylation, are termed "glycoforms." A cancer-specific glycoform is thus distinguishable from other (e.g., normal) glycoform(s) by a difference in glycosylation.

Differences in glycosylation that can produce distinct glycoforms include the removal of a glycan component, the addition of a glycan component, a change in the glycan component such as the substitution of one glycan component for another, and the rearrangement of one or more glycan components on the glycoprotein, as where a glycan component is shifted from one position on the polypeptide sequence to another.

Differences in glycosylation can be detected, according to the present invention, by utilizing a glycan-binding molecule, such as a glycospecific antibody, a glycospecific aptamer, or a lectin as further described herein, that is selective and/or specific for the cancer-specific glycoform. A particularly preferred glycospecific antibody is one that binds to an epitope that includes portions of both the polypeptide sequence and the glycan. Differences in glycosylation can also be detected spectroscopically. For example

mass spectrometry can be used to characterize the glycan component and distinguish glycoforms.

As used herein, the terms “cancer-specific glycoform” and “tumor-specific glycoform” are used interchangeably and refer to a glycoform of a glycoprotein which is found in a subject affected with cancer or a precancerous condition and which differs from a glycoform found in noncancerous tissue. The presence, absence or expression level of the cancer-specific glycoform is indicative of the presence, absence, nature, or extent of a cancer or a precancerous condition. In a cancer-specific glycoform, the protein component is the same as that found in the normal glycoform, but the glycan component is specific to a cancer or a precancerous condition. In other words, an exemplary cancer-specific glycoform is differentially glycosylated relative to the glycoform present in a normal or non-diseased sample. As shown in the following examples, a cancer-specific glycoform may be characterized by, for example, a GlcNAc $\beta(1,6)$ Man branched *N*-linked glycan component, a GlcNAc $\beta(1,4)$ Man bisected *N*-linked glycan component, an $\alpha(1,6)$ fucose linked to a core *N*-acetylglucosamine, or a branched *N*-linked glycan extended with *N*-acetylglucosamine. For example, a subject having cancer or a precancerous disease may express a glycoprotein containing a $\beta(1,6)$ branched *N*-linked glycan component while this $\beta(1,6)$ structure is absent from the analogous glycoprotein in a subject not having cancer or a precancerous condition. In this instance, the glycoprotein having the $\beta(1,6)$ branched *N*-linked glycan structure is the cancer-specific glycoform of that glycoprotein, and the detection of the cancer-specific glycoform is indicative of the presence of a cancer or a pre-cancerous condition.

Note that the designation $\beta(1,6)$ as in a $\beta(1,6)$ branched *N*-linked glycan is beta(1,6); the designation $\beta(1,4)$ as in $\beta(1,4)$ bisected *N*-linked glycan is beta(1,4); and the designation $\alpha(1,6)$ as in $\alpha(1,6)$ fucose linked to a core *N*-acetylglucosamine is alpha(1,6).

In some cases, the cancer-specific glycoform may be present in a disease-free individual, but expression of the cancer-specific glycoform may be elevated or reduced in an individual having cancer or a precancerous condition, in which case it is the change in expression level that is indicative of cancer or a precancerous condition. In other cases, the cancer-specific glycoform may be present only in a patient having cancer or a precancerous condition, and a different glycoform may be detectable in nonaffected individuals. In yet other cases, the glycoprotein itself (the cancer-specific biomarker), which in this case may exist only as a single glycoform may be found only in patients

having cancer or a precancerous condition, and may not be detectable in nonaffected individuals.

Cancer-specific glycoprotein glycoforms of the invention can be structurally characterized or identified by one or more of their protein sequence, their ability to bind
5 one or more glycan-binding molecules, and the chemical structure of their glycan component. An example of a cancer-specific glycoform is L-PHA reactive periostin, which can be identified or distinguished by, for example, sequential or concurrent contact with L-PHA lectin and an anti-periostin antibody. The ability of a cancer-specific glycoform to bind a particular glycan-binding molecule, e.g., a lectin, provides structural
10 information about the glycoform in that it indicates that the glycoform possesses one or more structural features, such as a particular glycosidic linkage, necessary to bind the glycan-binding molecule. One of skill in the art can, if desired, isolate and further characterize the glycoform, for example by performing a structural analysis to determine monosaccharide composition, linkage positions, branching, or sequence of the glycan
15 component, but such structural characterization of the glycan component is not necessary in order to practice the invention.

Examples of cancer-specific glycoprotein biomarkers that have been identified according to the identification method of the invention and which are particularly useful are described in more detail below and include, without limitation, periostin, osteoglycin
20 (also called mimecan), lysosomal-associated membrane glycoprotein 1 (LAMP-1), and lectin galactosidase soluble binding protein 3 (GALS3BP). Advantageously, cancer-specific glycoforms of these glycoproteins may be secreted or show residency on the outer surface of the cell from which they may be shed and find their way into the bloodstream or a bodily discharge.

25

Glycan-binding molecules

The present invention makes use of glycan-binding molecules to detect glycoprotein glycoforms. The term "glycan-binding molecule" refers to any molecule that is capable of binding to a glycan component of a glycoprotein. Preferably, the
30 glycan-binding molecule is glycoform-specific; that is, it selectively binds the glycan of one glycoform of a glycoprotein but not another, such that it can be used to distinguish different glycoforms of the glycoprotein. A glycoform-specific glycan-binding molecule is referred to herein as a "glycospecific." A glycan-binding molecule can be natural or synthetic. Examples of glycan-binding molecules include, without limitation, a lectin, a

glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, or a glycospecific small molecule. The term "aptamer" includes an RNA aptamer, a DNA aptamer, and a peptide aptamer, without limitation.

In a particularly preferred embodiment, the glycan-binding molecule selectively binds to a combination of the glycan component and the peptide component. For example, in the case of a glycan-binding molecule that is an antibody, the epitope includes portions of the glycan moiety as well as portions of the polypeptide sequence. This enhanced selectivity allows a single glycan-binding molecule to distinguish a glycoprotein from other proteins as well as from other glycoforms of the same protein. It also allows discrimination between glycoforms that have identical or essentially identical composition (i.e., same polypeptide sequence and glycan component) but wherein the glycan is attached at different locations in the polypeptide sequence as between the two glycoforms. In other words, the glycan-binding molecule can detect site-specific differences in glycosylation.

Preferably, the method of the invention utilizes a lectin to identify or detect a cancer-specific glycoprotein glycoform. Lectins are proteins or glycoproteins that bind to all or part of a glycan structure. Typically, lectins are non-enzymatic in action and are non-immune in origin. A lectin may bind to a glycan moiety which is part of a glycoprotein or another glycan-containing molecule such as glycolipids, glycoposphatidylinositols, and glycosaminoglycans. Lectins occur ubiquitously in nature and are found in both prokaryotes and eukaryotes, including bacteria, protozoa, fungi, plants and animals. Exemplary lectins include, without limitation, P-lectins, I-lectins, C-lectins, S-lectins (galectins), selectins, microbial carbohydrate proteins, glycosamineglycan binding proteins, and plant lectins. Lectins may also be produced synthetically by methods commonly used in the art such as recombinant DNA technology. A lectin useful in the method of the invention may be isolated from any source and may be naturally or synthetically produced, without limitation.

Importantly, lectins are capable of binding to specific glycans. Advantageously, the high specificity of a lectin for a particular glycan moiety enables the use of a lectin-glycoprotein binding to precipitate, isolate and/or detect glycoproteins (such as a cancer-specific glycoform of a glycoprotein) from or in a biological sample. For example, a type of N-linked glycosylation that is often increased in tumors, including breast and colon carcinoma, is the N-linked $\beta(1,6)$ branched glycan. This glycan structure associated with

the transition to malignancy is recognized by the lectin L-PHA, particularly when the glycan also expresses a distal $\beta(1,4)$ linked galactose.

Lectins and their specificities are widely known in the art (see Cummings and Etzler, "Antibodies and Lectins in Glycan Analysis," in *Essentials of Glycobiology* 5 *Second Edition*, Varki et al. (Eds.); Cold Spring Harbor Press: Woodbury, NY; 2009 and Tao et al., 2008 *Glycobiology* 18(10):761-769). Table 1 shows exemplary lectins and their binding specificities. It should be understood that Table 1 is illustrative only, and is not comprehensive with respect to either known lectins or with respect to the specificities shown for a particular lectin. Fig. 10 also shows exemplary lectin specificities.

10

Table 1. Examples of lectin specificities

Lectin	Abbreviation	glycan	schematic of glycan specificity* (determinants involved in binding are boxed)
<i>Aleuria aurantia</i> lectin	AAL	$\alpha(1,2)$, $(1,3)$, or $(1,6)$ -linked Fucose	
Concanavalin A (jack bean)	Con A	oligomannose-type N-glycan	
		hybrid-type N-glycan	
		biantennary complex-type N-glycan	
<i>Datura stramonium</i> lectin	DSL	tri-, tetraantennary complex-type N-glycan	
<i>Phaseolus vulgaris</i> lectin (red kidney bean; erythroagglutinin)	E-PHA	bisected di-, triantennary complex-type N-glycan	
<i>Phaseolus vulgaris</i> lectin (red kidney bean; leucoagglutinin)	L-PHA	tri-, tetraantennary complex-type N-glycan	
<i>Maackia amurensis</i> agglutinin	MAA	Neu5Ac α 2-3Gal	

Lectin	Abbreviation	glycan	schematic of glycan specificity* (determinants involved in binding are boxed)
<i>Sambucus nigra</i> agglutinin (Elderberry bark)	SNA	Neu5Acα2-6Gal or Neu5Acα2-6GalNAc	

* shading and shapes for the glycan structures reflect standard nomenclature adopted by the Consortium for Functional Glycomics. Specific structures shown and are as follows: Δ is fucose, $\circ$ is galactose, $\blacksquare$ is glucose, $\square$ is mannose.

** Information adapted from Cummings and Etzler, "Antibodies and Lectins in Glycan Analysis," in *Essentials of Glycobiology* Second Edition. Varki et al., (Eds.); Cold Spring Harbor Press: Woodbury, NY; 2009 and Tao et al., 2008 *Glycobiology* 18(10):761-769.

Generally, in the biomarker identification method of the invention, any lectin may be used, without limitation. In the diagnostic method of the invention, a lectin that is specific for the cancer-specific biomarker (e.g., glycoform) to be detected can be used.

Preferred lectins for use in the present invention include, but are not limited to, *Aleuria aurantia* lectin (AAL), concanavalin A (Con A), *Datura stramonium* lectin (DSL; also known as DHA, *Datura stramonium* agglutinin), *Phaseolus vulgaris* erythroagglutinin (E-PHA), *P. vulgaris* leucoagglutinin (L-PHA), *Maackia amurensis* agglutinin (MAA), and *Sambucus nigra* agglutinin (SNA, inclusive of SNAI and SNAII).

Identification of a biomarker

Included in the invention is a method for identifying a biomarker, preferably a cancer-specific glycoform of a glycoprotein.

In one embodiment of the method, a glycan-binding molecule, such as a lectin, is combined with a biological sample obtained from a patient with cancer or a precancerous condition, or from a cancer cell culture or animal with cancer or a precancerous condition (animal model), under conditions that allow the lectin to bind to a glycoprotein to form a lectin-glycoprotein complex. A control experiment is also performed. This lectin enrichment step coupled with a protein identification strategy identifies specific glycoproteins from the disease sample that are not enriched in the corresponding control sample. More particularly, the binding levels between the two samples are compared, and if they differ, the lectin-glycoprotein complex is optionally isolated, and the glycoprotein biomarker is identified. It should be noted that the presence, absence, increase in amount, decrease in amount, or differential glycosylation of the glycoprotein in the sample from the patient with cancer or a precancerous condition, compared to the analogous glycoprotein in the control sample, is indicative of the presence of a cancer-specific biomarker. The identification of cancer-specific biomarkers that selectively bind to a cancer-specific glycoform of the glycoprotein (thereby causing different binding levels

based on differential glycosylation of a glycoprotein found in both normal and disease samples) is particularly preferred.

Optionally, the glycan-binding molecule used in the method of identifying a biomarker may be tagged. The reasons and methods for tagging proteins are well known in art. For example, a protein may be tagged in order to facilitate isolation or tracking. Various types of tags that facilitate isolation or purification of the tagged biomolecule are commercially available and well known in the art and include, for example, beads (magnetic, sepharose, glass, agarose etc.), fusion peptides (hemagglutinin, 6-histidine, c-myc, fluorescent proteins, GST, etc.), or antibody-based technologies such as biotin-avidin. A tag may include a detectable marker. Detectable markers are widely used in the art and may be used for visualization. Examples of detectable markers include, but are not limited to, enzymatic reactions such as horseradish peroxidase, colorimetric readouts such as 3,3'-diaminobenzidine tetrahydrochloride (DAB), and the use of fluorescence, radioactivity, or chemiluminescence.

Methods for detecting proteins, and thereby methods for detecting lectin-glycoprotein binding, are also well known in the art include, for example, immunohistochemistry, immunocytochemistry, ELISA, immunoblotting (i.e. Western blotting). Methods of isolating a protein are well known in the art and include, for example, immunoprecipitation (IP), sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), or 2D gel electrophoresis.

The glycoproteins isolated in the lectin-glycoprotein complex may be identified by a number of different means that are commonly used in the art. Optionally, the components of the lectin-glycoprotein complex may be cleaved from each other prior to analysis through chemical or enzymatic cleavage. Preferably, the cleavage is enzymatic. Both the glycan and the protein components may be identified through methods that are routine in the field and well known in the art. Protein identification and sequencing methods are well known in the art and include, without limitation, Edman degradation and mass spectrometry. Typically, protein sequencing is followed by comparison of the protein sequence against databases such as *Entrez* (available at the National Center for Biotechnology Information website found on the world wide web at www.ncbi.nlm.nih.gov/) and UniProt (available at the European Bioinformatics Institute website found on the world wide web at www.ebi.ac.uk/uniprot/). These databases are publically available and the methods of using them are well known in the art.

--- Optionally, the method may include identifying or characterizing the glycan moiety present on the glycoprotein. The glycan component may be structurally identified by deglycosylating the isolated glycoprotein and subjecting the glycan component to methods commonly used to elucidate saccharide structures such as chemical analysis, mass spectrometry and high performance liquid chromatography (HPLC).

The method for identifying a biomarker optionally utilizes a panel of lectins. A “panel of lectins” is to be broadly understood to encompass one or more lectins, for example, at least 2 lectins, at least 5 lectins, or at least 10 lectins. The lectins used in the panel may have the same glycan specificity, similar glycan specificity or disparate specificity. Preferably, the panel of lectins consists of lectins having specificity for glycans containing branch points. More preferably, the panel of lectins consists of lectins having specificity for outer-branched glycans. A preferred panel of lectins contains at one, two, three, four, five, six or all seven lectins selected from AAL, ConA, DSL, E-PHA, L-PHA, MAA, and SNA, and optionally, additional lectins. A particularly preferred panel of lectins includes one or more of E-PHA, AAL, DSL, and/or L-PHA.

One of skill in the art will recognize that while the identification method is preferably performed using a lectin as the glycan-binding molecule, the method can be performed using any molecule capable of binding the glycan component of a glycoprotein, preferably a glycospecific molecule capable of selectively binding a glycoform of a protein. Such glycospecific molecules include, without limitation, a lectin, a glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, a glycospecific peptidomimetic or other small glycospecific molecule.

The biological sample used in the biomarker identification method of the invention is preferably a sample obtained from an organ or tissue, or a fluid associated with an organ or tissue, although it may also be a biological fluid such as blood or blood components, a bodily discharge, aspirate, and the like. More preferably, the biological sample of the method for identifying a biomarker is a tissue biopsy. The biological sample obtained from the patient having cancer or a precancerous condition preferably contains cancer cells. The sample may be pre-treated such that the cells are lysed, and partially purified to isolate, for example, a protein fraction or glycoprotein fraction prior to contact with the glycan-binding molecule. An example of an optional purification step is delipidation of the sample. The method therefore encompasses combining a glycan-binding molecule with one or more glycoproteins thus purified, partially purified or isolated from a biological sample (candidate biomarkers) to assay the particular

glycoprotein(s) for cancer-specificity. Differential binding of the glycoprotein(s) in the biological sample compared to a control sample is indicative of the presence of a candidate biomarker.

5 Material from an experimental model (including both animal models and culture systems) can also be used to identify a cancer-specific biomarker. In experimental models and cultures, carcinogenic properties or the cancer or precancerous condition may be spontaneously present or may be induced. One of skill in the art is familiar with methods for transforming cells and inducing tumors in experimental models.

10 A biological sample also includes *in vitro* culture constituents derived from organ, tissue, or cell culture including, but not limited to, conditioned media, tissue homogenates, whole cells, cell lysates, and cellular fractions or components. However, it is important to note that the cellular mechanisms and processes of *in vitro* cultures may not accurately represent the cellular mechanisms and processes of cells that have been isolated from an *in vivo* setting in which the cellular microenvironment plays an important
15 role in the maintenance of cellular mechanisms and processes, thus the use of biological samples obtained directly from a patient is preferable.

A control sample useful in the biomarker identification method of the invention includes a biological sample which is obtained from an experimental model (including both animal models and culture systems) that does not exhibit carcinogenic properties or
20 from a subject without cancer or a precancerous condition. The control sample may also be referred to as a non-diseased sample. Preferably, the control sample is obtained from the same source as the biological sample having cancer or a precancerous condition. For example, typically multiple biopsy samples are taken in parallel such that one sample contains tissue exhibiting cancer or a precancerous disease while a second or subsequent
25 sample contains adjacent, non-diseased tissue. Alternatively, tissue may be examined histologically and tissue exhibiting cancer or precancerous condition and non-diseased tissue present in the same tissue sample may be separated by techniques such as, but not limited to, laser capture microdissection. Control samples obtained from the same source are often referred to as tissue-matched controls. For example, when practicing the
30 biomarker identification method to identify breast cancer specific biomarkers, control samples are preferably obtained from the same patient using disease-free samples outside the tumor margin.

In some instances, a control sample may not be obtainable from the same source as the biological sample having exhibiting cancer or a precancerous condition. A control

sample may be obtained from a second model or subject that does not exhibit carcinogenic properties or from a subject without cancer or a precancerous condition.

Optionally, the control sample may be a pool of samples from multiple models that do not exhibit carcinogenic properties or from multiple subjects without cancer or precancerous conditions. For example, when practicing the biomarker identification method to identify ovarian cancer specific biomarkers, control samples (and optionally the experimental samples as well) are typically pooled samples.

In some embodiments, the control sample may be obtained from an experimental model or a subject exhibiting a particular disease state. A number of cancers are characterized by progressive disease states, such as those characterized by inflammation, dysplasia, hyperplasia, or cancer stages. In some tissues or organs, various disease states are distinct from one another. For example, a non-diseased pancreas may progress to either pancreatitis or to a pancreatic cancer. However, in other tissues or organs, there is a distinct disease progression. For example, non-diseased endometria may progress to endometrial hyperplasia which further progresses to endometrial cancer. In addition, non-diseased breast tissue may develop into breast hyperplasia that often progresses into a breast cancer. Thus, the method of the invention may be used to identify biomarkers specific to each stage of cancer or to specific disease states.

It will be appreciated that an enzyme that catalyzes a particular linkage specific to a cancer-specific glycoform of a glycoprotein (as compared with a normal glycoform) can also be cancer-specific biomarker. That is, once a cancer-specific glycoform of a glycoprotein has been identified, and the differential glycosylation has been characterized, enzymes such as glycosyltransferases (GT) and glycosylhydrolases (GH) that catalyze the cancer-specific glycosylation can also serve as cancer-specific biomarkers.

A number of cancer-specific biomarkers have been identified using the biomarker identification method of the invention. They include periostin, osteoglycin (also called mimecan), lysosomal-associated membrane glycoprotein 1 (LAMP-1), and lectin galactosidase soluble binding protein 3 (GALS3BP). Each of these biomarkers has a cancer-specific glycoform, which can be selectively detected using a lectin, as illustrated in the following Examples. Proteins involved in certain metabolic pathways are particularly preferred as breast cancer biomarkers, including those involved in urokinase plasminogen activator pathway and the TGF β pathway (such as the extracellular proteoglycans decorin and biglycan), as well as binding proteins related to cell adhesion,

cell-cell communication, organ development and metabolism, and proteins that respond to external stimulus.

Preferred glycoproteins for use as biomarkers specific to breast cancer include, but are not limited to, proteins listed in Table 4 (Example II). Breast cancer biomarkers that have glycoforms containing a $\beta(1,6)$ branched *N*-linked glycan component are preferred. Breast cancer biomarkers are preferably L-PHA reactive. In a preferred embodiment, a breast cancer biomarker is periostin, haptoglobin-related protein, apo-A1 precursor, osteoglycin, collagen VI $\alpha 3$, tubulin $\alpha 6$, collagen $\alpha 1$ VI chain precursor, collagen $\alpha 1$ VI isoform 2C2, variable Ig, triosphosphate iso, α -1 antitrypsin inhibitor, vimentin-like 50 kDa and 14-3-3- zeta protein. It is believed that neither osteoglycin nor 14-3-3- zeta protein have been previously associated with breast cancer, and further that this is the first report of a difference in glycosylation between the form of periostin found in normal breast tissue, and the form of periostin found in breast cancer tissue.

Preferred glycoproteins for use as biomarkers specific to ovarian cancer include, but are not limited to, proteins listed in Table 8 (Example IV). Ovarian cancer biomarkers that have glycoforms containing glycans that exhibit core fucosylation and bisected (also referred to as "bisecting") glycans, preferably bisected complex N-glycans, are preferred. Ovarian cancer biomarkers are preferably E-PHA reactive, AAL reactive, or DSL reactive. More preferably, ovarian cancer biomarkers may be selected from the group including periostin (a cancer-specific form is E-PHA reactive and possesses bisected N-linked glycans), biglycan, heparin sulfate proteoglycan 2, lactate dehydrogenase A, thrombospondin 1, serine protease inhibitor H1, lysosomal-associated membrane glycoprotein 1 (LAMP; both DSL-reactive and AAL-reactive forms are associated with ovarian cancer), and lectin galactosidase soluble binding protein 3.

Detection of cancers and precancerous diseases

The goal of cancer screening is to find cancers and precancerous conditions prior to the development of symptoms. As used herein the term "screening" refers to tests and examinations used to evaluate the presence, absence, nature or extent of a cancer or a precancerous condition. The term "early detection" refers to specific screening processes that allow detection and evaluation of a cancer or precancerous condition at an early point in disease progression. For example, early detection allows the evaluation of a cancer or a precancerous condition in subjects who do not yet display symptoms of the cancer or precancerous condition. Cancers or precancerous conditions that are detected due to the

manifestation of symptoms tend to be late-stage cancers that are relatively advanced and may have spread beyond the primary tumor site. In contrast, cancers found during early detection screens are more likely to be in an early stage and still be confined to the primary tumor site. It is widely understood and documented in the art that cancers that are detected early and remain as small, primary tumors are more easily treated and have better prognoses than cancers that are detected at later stages in which the tumors are large and have likely metastasized.

Diagnostic assay

Included in the present invention is a method for detecting a cancer or a pre-cancerous condition in a subject. The method utilizes the biomarkers of the invention, as described herein, and can be used alone or in combination with other procedures commonly used to detect or evaluate cancer or a precancerous condition. The method for detecting a cancer or a precancerous condition in a subject includes assaying for the presence of a biomarker and is useful for, among other things, diagnosing the presence of, evaluating the stage of, and determining prognosis of a cancer or precancerous condition.

In embodiments of the diagnostic assay wherein the biomarker is a glycoprotein glycoform, the diagnostic assay makes use of a glycan-binding molecule, such as a lectin, to discriminate among different glycans and thereby glycoforms, and optionally a second binding molecule, typically an antibody specific for the polypeptide component of the glycoprotein (regardless of isoform), to identify the glycoprotein. The antibody specific for the polypeptide component is referred to herein as a "protein antibody." For example, a cancer-specific glycoform of the cancer biomarker periostin can be recognized by the lectin L-PHA, but L-PHA is not specific for periostin, as it can recognize the same glycan on other glycoproteins. The L-PHA reactive periostin can be unequivocally identified as periostin through the use of a periostin antibody, which is commercially available from, for example, Abcam. Advantageously, the glycan-binding molecule and/or the protein antibody can be detectably labelled. A preferred detectable label is biotin.

In a preferred embodiment of the diagnostic method of the invention, the biomarker is a cancer-specific glycoform of a glycoprotein. Exemplary glycoprotein glycoforms are set forth elsewhere herein. It should be understood that the invention is generally applicable detection of any cancer-specific glycoform, without limitation. Cancer-specific glycoprotein glycoforms can include, for example, contain glycan components that include, for example, a GlcNAc β (1,6) Man branched *N*-linked glycan

component, a GlcNAc β (1,4) Man bisected *N*-linked glycan component, an α (1,6)-fucose linked to a core *N*-acetylglucosamine, or a branched *N*-linked glycan extended with *N*-acetyllactosamine.

The diagnostic assays of the invention are described below with reference to a
5 lectin, but it should be understood that any glycan-binding molecule, such as a
glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, or a
glycospecific small molecule, can be used in place of a lectin.

In one embodiment of the diagnostic method of the invention, a glycan-specific
molecule, such as a lectin, is combined with a biological sample isolated from a subject,
10 under conditions to allow lectin-glycan binding, yielding a lectin-glycoprotein complex.
The resulting lectin-glycoprotein complex is isolated, and the lectin-reactive glycoprotein
biomarker is detected, wherein detection of the biomarker is indicative of presence of
cancer or a precancerous condition. Methods of detecting or measuring the levels of a
protein, including a glycoprotein, are well known in the art and include, without
15 limitation, immunoprecipitation, immunohistochemistry, immunocytochemistry, ELISA,
and immunoblotting (i.e. Western blotting). The glycoprotein biomarker can be
conveniently detected using a detectably labelled antibody specific for the protein (a
protein antibody). Optionally, the method may further include comparing the amount of
biomarker to the amount of biomarker in a reference sample, or to a reference level, to
20 determine whether the amount of the biomarker is indicative of the presence of cancer or
a precancerous condition.

Any biological sample can be tested, without limitation. Advantageously, the
invention permits detection of the biomarker in bodily fluids, providing a convenient,
low-cost screening option. Biological samples that can be tested in the diagnostic method
25 of the invention include, without limitation, organs, tissues (including biopsies), fecal
matter, bone marrow, lymph tissue, biological fluids, and bodily discharges obtained from
a human or veterinary subject. Biological fluids may include, without being limited to,
blood (including components of blood such as serum or plasma), urine, bile, spinal fluid,
lymph fluid, ascites fluid, pancreatic ductal fluid, sputum, pleural fluid, tears, saliva,
30 mucus, breast milk and bodily discharges (such as vaginal discharge, nasal discharge, and
nipple aspirate). A biological fluid is also intended to include breath in both a vaporous
and liquid form. Preferably, the biological sample is obtained from a subject having
cancer or a precancerous condition, although the screening method of the invention
contemplates that both diseased and nondiseased subjects will be screened. Preferably,

the biological sample in the diagnostic method of the invention is a biological fluid. More preferably, the biological sample is serum, vaginal discharge, or nipple aspirate. Even more preferably, the biological sample is serum.

In another embodiment, the diagnostic method includes combining a glycoform-specific lectin with one or more purified or partially purified glycoproteins obtained from a biological sample, enabling lectin-glycan binding to form a lectin-glycoprotein complex, and detecting a cancer-specific glycoform of the glycoprotein, wherein detection of the cancer-specific glycoform is indicative of presence of cancer or a precancerous condition. Isolation or purification of the glycoprotein(s) prior to contact with the lectin can be accomplished by fractionating the sample using any convenient method, for example by contacting the sample with a lectin with lesser specificity (e.g., concanavalin A, which binds high mannose, hybrid, and complex biantennary glycans). Likewise, the biological sample can be first contacted with a protein antibody to isolate the glycoprotein, then contacted with the glycoform-specific molecule, such as a lectin, to determine whether the isolated glycoprotein is a cancer-specific glycoform of the glycoprotein. Optionally, the method may further include comparing the amount of biomarker to a reference, wherein altered glycosylation or altered expression of the biomarker relative to the reference is indicative of the presence of cancer or a precancerous condition.

In a particularly preferred embodiment of the diagnostic method of the invention, the glycan-binding molecule selectively binds to a combination of the glycan component and the peptide component. For example, in the case of a glycan-binding molecule that is an antibody, the epitope includes portions of the glycan moiety as well as portions of the polypeptide sequence. This enhanced selectivity allows a single glycan-binding molecule to distinguish a glycoprotein from other proteins as well as from other glycoforms of the same protein, thereby performing in a single step the functions of the glycan discrimination and protein identification.

Furthermore, it should be understood that detection of a cancer-specific glycoform or other biomarker in the diagnostic assay of the invention is not limited to chemical or immunochemical binding methods but includes any method of detection, without limitation, including for example spectroscopic detection such as mass spectrometric analysis, fluorescence, magnetic, electromagnetic or optical methods, or chemical analysis.

In another embodiment of the diagnostic method of the invention, a subject's bodily fluids, tissues or organs can be assayed for the presence of an autoimmune response to cancer or a precancerous condition. More specifically, the subject can be screened for the presence of circulating antibodies to one or more cancer biomarkers of the invention. For example, antigen that include a peptide sequence and glycan component that characterize a cancer-specific glycoform of a glycoprotein can be synthesized, and the subject's bodily fluid, tissue or organ can be contacted with the antigen to determine the presence of an antibody that binds thereto. The diagnostic assay is readily scalable and amenable to multiplexing in order to facilitate cancer screening using a multiplicity of synthetic cancer biomarkers.

In embodiments of the diagnostic assay wherein the biomarker is a glycosyltransferase (GT) or a glycosylhydrolyase (GH) enzyme that catalyzes the formation of a cancer-specific glycoprotein glycoform, the diagnostic assay typically measures RNA transcript levels for the enzyme, or enzyme activity.

Accordingly, in another embodiment, the diagnostic method involved detecting an enzyme biomarker that catalyzes a particular linkage characterizing a cancer-specific glycoform of a glycoprotein. Detection of an enzyme may occur at either a protein or a transcript level. Methods of detecting a protein include, without limitation, immunoprecipitation, immunohistochemistry, immunocytochemistry, ELISA, and immunoblotting (i.e. Western blotting). Examples of assays that may be used to detect a transcript are well known in the art and include, for example, Northern blot assays, reverse transcription polymerase chain reactions, RNase protection assays, and the like. Optionally, the method for detecting a biomarker in a subject, wherein the biomarker is an enzyme that catalyzes the particular linkage characterizing the cancer-specific glycoform of a glycoprotein may further include comparing the amount of biomarker to a reference, wherein altered expression of the biomarker is indicative of the presence of cancer or a precancerous condition.

Preferred enzyme biomarkers include, without limitation, enzymes encoded by the genes listed in Table 6 (Example III). Particularly preferred biomarkers include, without limitation N-acetyl glucosaminyltransferase V (GnT-V), fucosyltransferase 8 (FUT8), mannosidase N-acetylglucosaminyltransferase 3 (MGAT3), MGAT4a, MGAT4b, MGAT5, and MGAT5b. GnT-V catalyzes the addition of $\beta(1,6)$ branched *N*-linked glycans. FUT8 catalyzes the addition of core fucosylated *N*-linked glycans. MGAT3 catalyzes the addition of bisected *N*-linked glycans. MGAT4a and MGAT4b catalyze the

addition of outer-branched *N*-linked $\beta(1,4)$ glycans. MGAT5 catalyzes the addition of outer branched *N*-linked $\beta(1,4)$ glycans. These glycan moieties are recognized by preferred lectins of the invention such as AAL, ConA, DSL, E-PHA, L-PHA, MAA, and SNA.

5 A “reference” includes, without limitation, a control sample (as previously defined) that may be used in comparison against the biological sample isolated from a subject. For example, the reference may be obtained from a subject without cancer or a precancerous condition. Optionally, the reference may be a pool of samples obtained from at least two subjects without cancer or a precancerous condition. Alternatively, the
10 reference may be obtained from a subject having a cancer or a pre-cancerous condition at a determined stage. Alternatively, the reference may be a previously obtained biological sample from the subject. Alternatively, the reference may be a published or commonly known level of the biomarker (Galen, *Beyond Normality: the predictive value and efficiency of medical diagnosis*, Wiley & Sons: New York, NY; 1975). For example,
15 prostatic states such as non-diseased prostate, benign prostatic hyperplasia (BPH), and prostate cancer each exhibit a particular range of the prostate specific antigen (PSA) biomarker and the level of PSA associated with each state is commonly accepted by clinicians and used as a reference in diagnostic assays.

 As used herein, the terms “cancer” and “pre-cancerous condition” refer to any
20 uncontrolled and/or undesired growth of cells. The cancer or pre-cancerous condition is intended to include a hyperplastic growth, a benign tumor, a malignant tumor, or a metastasized tumor. A tumor is not limited to a solid tumor, but is intended to include any uncontrolled and/or undesired growth of cells including, for example, blood tumors that do not form a solid mass. The tumor may consist of, without being limited to,
25 epithelial cells, stromal cells, undifferentiated cells, or any combination thereof. An epithelial cell is any cell that covers a surface, or lines a cavity or the like, and that, in addition, performs the functional aspect of the tissue such as any secretory, transporting, or regulatory function. Epithelial cells may be further classified as, for example, endothelial or mesothelial cells. Nonlimiting examples of epithelial cells include
30 squamous, cuboidal, columnar, transitional, simple, stratified, and secretory. A stromal cell, also referred to as a mesenchymal cell, provides support for or surrounds tissues and organs. Nonlimiting examples of stromal cells include fibroblasts, immune cells, pericytes, endothelial cells and inflammatory cells. A carcinoma is a tumor derived from or consisting primarily of epithelial cells. Preferably, the cancer or pre-cancerous

condition of the present invention is a carcinoma. Exemplary carcinomas for use in the present invention include, but are not limited to, breast, ovarian, colon, rectal, colorectal, pancreatic, or liver carcinomas.

5 This diagnostic method of the invention is highly amenable to multiplexing. For example, a panel of glycoform-specific (glycospecific) binding molecules (binding elements), such as lectins, glycospecific antibodies, or glycospecific aptamers, can be used to assay the biological sample for the presence or absence of cancer-specific glycoform biomarkers. The glycoform-specific binding molecules selected for use in the multiplexed assay preferably, but need not, bind cancer-specific glycoforms of a plurality
10 of different glycoproteins, e.g., periostin and osteoglycin. The glycoproteins present in the glycoprotein/lectin complexes can then be identified using protein-specific antibodies. In another example, a panel of protein-specific antibodies (binding elements) can be used to assay the biological sample for the presence or absence of different glycoproteins that are known to possess a cancer-specific glycoform, then the bound glycoprotein/antibody
15 complexes can be further assayed for the presence of cancer-specific glycoforms by contacting them with a plurality of glycospecific molecules such as lectins, glycospecific aptamers, or glycospecific second antibodies known to selectively bind the cancer-specific glycoforms.

A multiplexed panel is to be broadly understood to encompass a multiplicity of
20 binding elements. A multiplexed panel can include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more biomarkers. Preferably, the biomarker panel includes at least two biomarkers selected from periostin, osteoglycin, lysosomal-associated membrane glycoprotein 1, and lectin galactosidase soluble binding protein 3. More preferably, the biomarker panel includes at least periostin and osteoglycin.

25 Preferably, the method for identifying a biomarker is a high-throughput assay. High throughput assays allow numerous samples to be reviewed rapidly and simultaneously for the presence of one or more particular glycan components. Typically, a high-throughput assay is performed in a micro-titer plate that enables the review of, for example, at least 6, at least 24, at least 96, at least 384, at least 1536, or at least 3456
30 samples. Assays that use micro-titer plates are well known in the art and are capable of, for example, detecting molecular interactions, detecting cell growth, detecting enzymatic activity, nucleic acid quantitation, and immunoassays. Detection methods commonly used in such assays include, but are not limited to absorbance, fluorescence intensity, and luminescence. Preferably, the method of identifying a biomarker utilizes enzyme-linked

immunosorbent assay (ELISA) to detect lectin-glycoprotein binding. Alternatively, a high-throughput assay may include any assay that simultaneously evaluates a large number or samples, such as microarray technologies.

5 *Therapeutic applications*

Also included in the invention is a therapeutic method and pharmaceutical composition for treating a subject having cancer or a precancerous condition. The method and pharmaceutical composition for treating a subject having cancer or a precancerous condition can be used alone or in combination with other procedures
10 commonly used to treat cancer or a precancerous condition such as chemotherapy, surgery, and radiation therapy.

Cancer-specific glycoforms contain aberrant glycan moieties relative to the glycoform present in a non-diseased subject. Glycosylation and the resulting glycan moiety affect the ability of the protein to transduce normal cellular signals. For example,
15 the glycan moiety present on a cancer-specific glycoform may cause a physical block of a typical binding interaction, prevent typical migration, induce atypical migration, or otherwise interfere with the normal signal transduction of the glycoprotein to result in mis-regulated cellular signaling that can result in the development of cancer or a precancerous condition. Therefore, inhibiting the production of or the function of a
20 cancer-specific glycoform may restore normal signal transduction pathways to treat or prevent cancer or a precancerous disease.

In one embodiment, the method for treating a subject having cancer or a precancerous condition may include administering a therapeutic agent capable of targeting a biomarker, wherein the biomarker is a cancer-specific glycoform of a
25 glycoprotein. Administration of the therapeutic agent can be prophylactic or, alternatively, can be initiated after the development of a cancer or a precancerous condition. Treatment that is prophylactic, for instance, initiated before a subject manifests symptoms of a cancer or a precancerous condition, is referred to herein as treatment of a subject that is "at risk" of developing the condition. Treatment initiated
30 after the development of a cancer or a precancerous condition may result in decreasing the severity of the symptoms of one of the conditions, or completely removing the symptoms.

The term "therapeutic agent," as used herein, refers to a molecule capable of inhibiting a cancer-specific glycoform of the glycoprotein. Inhibition may occur at a

number of different levels. For example, one could prevent transcription, prevent translation, prevent posttranslational modification, or prevent function of the cancer-specific glycoform. Methods for preventing transcription are well known in the art and include, without limitation, and short hairpin RNAs (shRNAs), morpholinos, and anti-sense DNA-hybridizing probes. Methods for preventing translation are well known in the art and include, without limitation, RNA interference (RNAi), and small interfering RNAs (siRNA). Methods for preventing posttranslational modification include inhibiting the enzyme that catalyzes the particular linkage characterizing the cancer-specific glycoform by any of the inhibition methods listed herein. Methods for preventing the function of the cancer-specific glycoform are well known in the art and include, without limitation, antibody technology, peptidomimetics, and small molecule inhibitors. Preferably, the therapeutic agent of the invention prevents the aberrant glycosylation of the cancer-specific glycoform of a glycoprotein.

Also included in the invention is a composition containing the therapeutic agent. A composition may be prepared by methods well known in the art of pharmacy. In general, a composition can be formulated to be compatible with its intended route of administration. Typically, the composition further contains a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable," as used herein, means that the therapeutic agent so described is suitable for use in contact with a subject without undue toxicity, incompatibility, instability, allergic response, and the like. Typically, pharmaceutically acceptable carriers include saline, solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration.

The therapeutic agent may be administered by a number of methods including, but not limited to, oral and systemic administration. Therapeutic amounts are amounts which eliminate or reduce the patient's tumor burden, or which prevent or reduce the proliferation of metastatic cells. Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the therapeutic agent can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. For administration by inhalation, the active compounds are delivered in the form of an aerosol spray from a pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. Systemic administration includes transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used

in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

Preferably the therapeutic agent is a glycan-binding molecule including, but not limited to lectins, glycospecific antibodies, and any small molecules capable of specifically binding to the cancer-specific glycoform. More preferably, the therapeutic agent is a lectin or a glycan-specific antibody.

Optionally, the method may further include conjugating the therapeutic agent to a cytotoxic agent, further enhancing toxicity to targeted cells. Such agents include, but are not limited to, known chemotherapeutic agents, immunotherapeutic agents, and radiotherapeutic agents.

In another embodiment, the invention may include a method for treating a subject having cancer or a precancerous condition, wherein the glycan moiety present on the cancer-specific glycoform of the glycoprotein is a targeting moiety. Glycans are typically maintained on the cell surface. Therefore, the cancer-specific glycoform of a glycoprotein provides a novel target for delivering therapeutic agents to the tumor. Currently, one of the major issues in cancer therapies is effectively targeting the tumor cells. Consequently, a majority of the cancer therapies currently available are toxic to many healthy cells as well as to cancer cells.

The method for treating a subject having cancer or a precancerous condition, wherein the glycan moiety present on the cancer-specific glycoform of the glycoprotein is a targeting moiety may include administering a glycan-binding molecule conjugated to a therapeutic agent to a subject having cancer or a precancerous condition, wherein the glycan-binding molecule or glycopeptide binding specifically targets the glycan moiety on the cancer-specific glycoform of the glycoprotein.

It should be noted that although the invention is described primarily with respect to cancer and precancerous conditions in humans, it is equally applicable to all mammalian subjects and, in that regard, has application in veterinary and research settings as well as in human medical contexts. Preferably, the subject is a mammal. A mammal may include, without limitation, domestic animals (such as cats, dogs, horses, and cattle), laboratory animals (such as mice, rats, rabbits, and monkeys), and humans. More preferably, the subject is a human.

The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the invention as set forth herein.

EXAMPLES

Example I

Identification of Glycoproteins Associated with Invasive Human Breast Cancer Using Lectin Glycoproteomics

Glycosylation is a dynamic post-translational modification that changes during the development and progression of various malignancies. Oncogenically transformed cells demonstrate defined changes in glycan structures, in particular they show increased amount of a specific type of N-linked structure known as the $\beta(1,6)$ branch (Pierce et al., 1997 *Glycoconj J* 14:623-630). The enzyme known as N-acetylglucosaminyltransferase V (GnT-V) adds GlcNAc in a $\beta(1,6)$ linkage to form the branch that leads to the formation of complex polylactosamine structures. During the oncogenesis of breast carcinoma, GnT-V transcript levels and activity are increased due to activated oncogenic signaling pathways. Elevated GnT-V levels leads to increased $\beta(1,6)$ branched N-linked glycan structures on glycoproteins that can be measured by the binding of a specific lectin, L-PHA. L-PHA immunohistochemical staining on paraffin sections of normal breast carcinoma, ductal hyperplasia, and carcinoma in situ reveal that L-PHA staining was qualitative at the point cells are progressing to carcinoma in situ (Fernandes et al., *Cancer Res.*, 1991. 51:718-723). The L-PHA staining of breast carcinoma frequently shows a coarse granular type punctate staining pattern. L-PHA does not bind to non-diseased breast epithelial cells, but during the progression to invasive carcinoma, cells show a progressive increase in L-PHA binding.

We have developed a procedure for intact protein L-PHA-affinity enrichment followed by electrospray ionization mass spectrometry (ESI-MS) to identify potential biomarkers for the early detection of breast carcinoma. We have characterized the media

from the invasive breast cancer cell line MBAMB231 for L-PHA-reactive proteins, and found that this methodology clearly detects glycoproteins with $\beta(1,6)$ branched glycan structures as well as proteins that are associated with these glycoproteins. Our results identified proteins participating in the urokinase plasminogen activator pathway and the
5 TGF β pathway.

In example II we analyzed matched normal and malignant tissue from patients with invasive breast ductal carcinoma. L-PHA enrichment coupled with ESI-MS identified specific glycoproteins from the tumor tissue that are not enriched in the corresponding normal tissue sample for the same patient. These results indicate that
10 lectin enrichment strategies that target glycan changes known to occur during the oncogenesis of a particular tumor can be a powerful method of biomarker discovery.

Methods

Cells were cultured in serum-free media. Proteins secreted from cultured cells
15 were collected from the media, concentrated, and the buffer was exchanged. Biotinylated L-PHA lectin was added to the proteins. Magnetic streptavidin particles were then added to the solution containing both the proteins and the biotinylated L-PHA. Lectin-bound glycoproteins were captured using a magnetic stand, washed, and eluted in urea/DTT. The eluted proteins were carboxyamidomethylated and subjected to a tryptic digestion
20 and C18 reverse phase chromatography.

CID fragmentation was carried out on an LTQ. All experiments were performed using nanospray of eluent from a reverse-phase C18 capillary column (75 μ m by 10 cm) using an increasing gradient of acetonitrile in formic acid. Data was stringently filtered using Turbo Sequest to achieve a false positive rate of less than 1.5 % for any single
25 peptide (protein false positive rate of less than 0.3 % with 2+ peptides) using an inverted database.

Results

Defining the specificity of L-PHA binding. GnT-V is the only glycosyltransferase
30 in breast cancer cells that adds the branched $\beta(1,6)$ glycan structure required for L-PHA binding. Therefore to test the specificity of the L-PHA lectin we have suppressed the expression of GnT-V using RNA interference. Fig. 1 is a lectin blot showing the reactivity of L-PHA toward total cellular proteins extracted from wild-type and GnT-V(-)

MDA-MB231 invasive breast carcinoma cells (the three bold bands are non-specific bands resulting from the streptavidin-horseradish peroxidase detection).

L-PHA affinity enriched secreted glycoproteins identified by ESI-MS from MDA-MB231 cells. Glycoproteins that were bound by L-PHA and isolated from the serum-free media of cultured cells are shown in Table 2.

Table 2. L-PHA affinity enriched secreted glycoproteins from serum-free media of cultured cells.

Protein	Function	presence of secretion signal
A4 amyloid protein precursor	binds cationic trypsinogen to inhibit activity	yes
alpha actinin	actin binding protein promotes cell migration	possible
alpha enolase	glycolytic enzyme can serve as a plasminogen receptor	yes
annexin A1	part of plasminogen activator system	yes
annexin A2	part of plasminogen activator system	yes
annexin A5	part of plasminogen activator system	yes
ARCNI protein	coatamer protein	yes
ATP citrate lyase isoform 2	cell growth	no
cathepsin D	protease found in lamellar granules	yes
cationic trypsinogen	cell proliferation	yes
folistatin-related protein	regulator of activin and TGF-beta signaling	yes
galectin 3 binding protein	tumor antigen associated with metastasis	yes
lysosomal membrane glycoprotein-2 (lamp-2)	adhesive glycoprotein, ligand of galectin-3 promotes tumor cell growth and invasion	yes
neonatal thrombolytic agent	plasminogen activator promotes tumor metastasis	no
PAI precursor	promotes tumor cell invasion by degrading extracellular matrix-part of the plasminogen activator system	yes
pyruvate kinase 3 isoform 2	glycolytic enzyme	yes
S100 calcium binding protein	calcium dep membrane binding protein interacts with annexin proteins	

Protein	Function	presence of secretion signal
tenascin C	promotes tumor growth and angiogenesis	yes
thrombospondin 1 precursor	multifunctional matrix glycoprotein influences tumor growth	yes
TRAP1 (TNF-1 receptor associated)	chaperone activity activation of TNF alpha	yes
vimentin	cytoskeletal protein found in mesenchymal tissue mainly, except in migrating endothelial cells	no

Conclusions

GnT-V glycosylation correlates with transition from normal breast epithelium to carcinoma in situ. GnT-V glycosylation can be identified by the specific and high affinity binding of the lectin L-PHA. In this evaluation of secreted L-PHA reactive glycoproteins from an invasive breast carcinoma cell line we have specifically captured, identified, and initiated mapping of N-glycosylated peptides using mass spectrometry. Our results have identified several glycoproteins implicated in cell proliferation and cell migration control mechanisms. In Example II we show the use of this lectin affinity method to pull out $\beta(1,6)$ branched glycoproteins from matched normal and ductal invasive carcinoma tissue samples, correlate the results and identify potential biomarkers for the development of a serum-based breast cancer screening assay as illustrated in Example V.

Example II

Targeted Glycoproteomic Identification of Biomarkers for Human Breast Carcinoma

Glycosylation is clearly the most complex set of post-translational modifications that proteins undergo during biosynthesis, and several specific types of glycan epitopes have been shown to be associated with various types of cancer (Kim and Varki, 1997. *Glycoconj J* 14:569-76; Hakomori, 2001. *Adv Exp Med Biol* 491:369-402). The glycosylation patterns of cell surface glycoproteins play important roles in mediating cell-cell and cell-matrix interactions. During oncogenesis, distinct signal transduction pathways are altered, leading to the differential expression of numerous genes. Genes known as glycosyltransferases (GT) and glycosylhydrolases (GH), responsible for the

addition and removal of sugars on proteins in the ER and Golgi apparatus, can change activity during oncogenesis, causing different oligosaccharide structures to emerge on cell surface glycoproteins (Pierce and Arango, 1986. *J. Biol. Chem.* 261:10772–10777; Meezan et al., 1969. *Biochemistry* 8:2518–2524). These glycan changes can have potent effects on the tumor microenvironment, promoting tumor invasion and metastasis (Guo et al., 2002. *Cancer Res.* 62:6837–6845; Guo et al., 2003. *J. Biol. Chem.* 278:52412–52424; Guo et al., 2007. *J. Biol. Chem.* 282:22150–22162; Abbott et al., 2006. *Exp. Cell Res.* 312:2837–2850). Factors contributing to the regulation of glycosylation include: nucleotide sugar donor availability, substrate availability, sequential reactions, and transcriptional regulation of GT and GH. Studies examining the dynamics of the glycome during differentiation of stem cells have found that changes in GT and GH mRNA levels correlate well with glycan structures observed (Nairn et al., 2007. *J. Proteome Res.* 6:4374–4387). Therefore, despite the complexity of factors influencing glycosylation, transcriptional regulation of the enzymes involved in the synthesis and catabolism of glycans seems to be the one of the primary mechanisms to control glycan structures on the cell surface. Comparative studies examining the differences in GT and GH expression patterns between normal and tumor tissue could direct the discovery of tumor-specific glycosylation changes associated with particular malignancies, which could then be exploited to develop diagnostic and cell targeting reagents.

A particular type of N-linked glycosylation that is often increased in tumors is the N-linked $\beta(1,6)$ branched glycan (Pierce and Arango, 1986. *J Biol Chem* 261:10772-7) that is bound by the lectin, L-PHA, when the glycan also expresses a distal $\beta(1,4)$ linked galactose (circled in Fig. 2; Cummings and Kornfeld, 1982. *J Biol Chem* 257:11235-40; Cummings and Kornfeld, 1982. *J Biol Chem* 257:11230-4). For example, staining of normal breast epithelia shows insignificant reactivity with L-PHA; yet, in breast carcinoma, staining by this lectin significantly increases (Fernandes et al., 1991. *Cancer Res* 51:718-23; Dennis and Laferte, 1989. *Cancer Res* 49:945-50). Expression of $\beta(1,6)$ branched glycan structures in both breast and colon carcinoma appears to be a qualitative change exhibited at the transition to malignancy, which is likely caused by up-regulation of the glycosyltransferase known as GnT-V (GnT-Va, Mgat5a) that synthesizes the N-linked $\beta(1,6)$ branch (Buckhaults et al., 1997. *J Biol Chem* 272:19575-81). This glycosylation is often a step toward the formation of more complex poly N-acetyllactosamine glycan structures (Fig. 2) which serve as ligands for the class of animal

lectin known as galectins that are often elevated in metastatic carcinoma (Lagana et al., 2006. *Mol Cell Biol* 26:3181-93). A recent study staining more than 700 primary breast tumors with L-PHA found that $\beta(1,6)$ branched oligosaccharides were an independent prognostic indicator for poor outcome in primary node-negative tumors (Handerson et al., 5 2005. *Clin Cancer Res* 11:2969-73).

Experimental modulation of GnT-V activity, both in vivo and in vitro, results in changes in carcinoma invasiveness and metastasis, supporting the conclusion that increases in the posttranslational modification of proteins by GnT-V is a mechanism by which tumor cell malignancy may increase (Guo et al., 2002. *Cancer Res* 62:6837-45; 10 Guo et al., 2003. *J Biol Chem* 278:52412-24; Cheung et al., 2007. *Glycobiology* 17:828-37; Guo et al., 2007. *J Biol Chem* 282:22150-62).

In this example, we show the development of a simple glycoproteomic strategy to identify the glycoproteins from breast tissue that bind to a specific carbohydrate binding protein or lectin known as L-PHA (L-phytohemagglutinin). Glycosylation is a dynamic 15 post-translational modification that changes during the development and progression of various malignancies. During the oncogenesis of breast carcinoma, the glycosyltransferase known as N-acetylglucosaminyltransferase Va (GnT-Va) transcript levels and activity are increased due to activated oncogenic signaling pathways. Elevated GnT-V levels leads to increased $\beta(1,6)$ branched N-linked glycan structures on 20 glycoproteins that can be measured using L-PHA. L-PHA does not bind to non-diseased breast epithelial cells, but during the progression to invasive carcinoma, cells show a progressive increase in L-PHA binding. Our experimental design utilizes intact proteins and allows for the identification of glycoproteins whose glycans are bound by the lectin L-PHA, as well as those proteins/glycoproteins that may associate with the bound 25 glycoproteins (Abbott et al., April 2008 *J. Prot. Res.* 7(4):1470-1480; Abbott et al., April 2008 *J. Prot. Res.* 7(4):1470-1480 online Supporting Information available at http://pubs.acs.org/doi/suppl/10.1021/pr700792g/suppl_file/pr700792g-file004.pdf).

Here we describe a procedure for intact protein L-PHA-affinity enrichment, followed by nanospray ionization mass spectrometry (NSIMS/MS), and subsequent 30 proteomic data analysis to identify potential biomarkers for breast carcinoma. Our results demonstrate that this technique can be an effective method to identify proteins with tumor-specific glycosylation changes. We identified L-PHA reactive glycoproteins from matched normal (non-diseased) and malignant tissue isolated from patients with invasive ductal breast carcinoma. Comparison analysis of the data identified 34 proteins that were

enriched by L-PHA fractionation in tumor relative to normal for at least 2 cases of ductal invasive breast carcinoma (Table 4). Of these 34 L-PHA tumor enriched proteins, 12 (periostin, haptoglobin-related protein, apo-A1 precursor, osteoglycin, collagen VI α 3, tubulin α 6, collagen α 1 VI chain precursor, collagen α 1 VI isoform 2C2, variable Ig, triosphosphate iso, α -1 antitrypsin inhibitor and vimentin-like 50 kDa) are common to all 4 matched cases analyzed. These results indicate that lectin enrichment strategies targeting a particular glycan change associated with malignancy can be an effective method of identifying potential biomarkers for breast carcinomas with diverse clinical features.

Materials and Methods

Specimens. Tissue specimens, matched normal and tumor, from patients with histologically proven invasive ductal breast carcinoma were collected in accordance with approved institutional review board Human subject's guidelines at Emory University Hospital, Atlanta, GA. Board certified clinical oncologists and pathologists carried out all clinical and histological analysis of the biopsies. All specimens for this study were immediately frozen at -70°C until proteomic analysis. For the initial validation of our L-PHA affinity enrichment method we analyzed 4 patients with matched normal and malignant breast tissue (Table 3).

Sample processing, L-PHA enrichment, and MS. Frozen tissue samples were processed as follows: 100 mg of tissue was de-lipidated using a mixture of chloroform/methanol/ water (4:8:3, v/v/v) as described previously (Seppo et al., 2000. *Eur J Biochem* 267:3549-58; Aoki et al., 2007. *J Biol Chem* 282:9127-42). Delipidated and precipitated proteins were pelleted by centrifugation and the pellet was given an additional wash with acetone and water (4:1) on ice for 15 minutes. Intact proteins were extracted from the de-lipidated tissues using a mild detergent solution as follows: 5 mg of delipidated protein powder was dissolved in 300 μl of 50 mM Tris-Cl pH 7.5, 0.1 % NP40, 150 mM NaCl, 0.4 mM EDTA, 1 protease inhibitor tablet, the sample was sonicated 3 times for 10 second pulses at setting 5 (Vertis Virsonic microtip). The supernatant was taken after centrifugation at 10,000 rpm at 4°C for 10 minutes. The protein concentration of the sample was determined by BCA assay and 600 μg of total protein lysate was dialyzed overnight at 4°C into 40 mM ammonium bicarbonate using a 4,000 MWCO tube-O-dialyzer (GBiosciences). The sample was adjusted to 150 mM NaCl, 5 mM CaCl_2 , and 5 mM MgCl_2 before the addition of the lectin. Biotinylated L-PHA (Vector Labs,

Burlingame, CA) (10 µg) was added and the sample was rotated at 4°C overnight. Bound L-PHA reactive proteins were captured using 100 µl paramagnetic streptavidin particles (Promega) at 4°C for 2 hours. After extensive washing in 1X PBS, captured proteins were eluted with 200 µl of 2M Urea/0.2 mM DTT/40 mM ammonium bicarbonate at 52°C for 1 hour. The eluted fraction was separated from the paramagnetic streptavidin particles using a magnetic stand. Eluted proteins were carboxyamidomethylated by adding an equal volume of iodoacetamide (10 mg/ml in 40 mM ammonium bicarbonate) in the dark for 45 minutes and digested with 5 µg of sequencing grade trypsin (Promega) at 37°C overnight. Tryptic peptides were acidified with 200 µl of 1 % trifluoroacetic acid and desalting was performed using C18 spin columns (Vydac Silica C18, The Nest Group, Inc.). Eluted peptides were dried in the speed vac and resuspended in 78 µl buffer A (0.1 % formic acid) and 2 µl of buffer B (80 % acetonitrile/0.1 % formic acid) and filtered through a 0.2 µm filter (nanosep, PALL). Samples were loaded off-line onto a nanospray column/emitter (75µm x 8.5 cm, New Objective) self-packed with C18 reverse-phase resin in a nitrogen pressure bomb for 10 minutes. Peptides were eluted via a 160-minute linear gradient of increasing B at a flow rate of approximately 250 nl/min. directly into a linear ion trap (LTQ, Thermo Co. San Jose, CA equipped with a nanoelectrospray ion source). The top eight ions from the full MS (300-2000M/Z) were selected for CID fragmentation at 34 % with a dynamic exclusion of 2.

Permethylation of Glycans. To facilitate the analysis of oligosaccharides by MS, N-linked glycans released by N-glycanase were permethylated as described previously (Aoki et al., 2007. *J Biol Chem* 282:9127-42). Briefly, following extraction from tissue samples, delipidated proteins were digested with trypsin and chymotrypsin. The resulting digests were enriched for glycopeptides, which were then treated with PNGaseF (Prozyme) to release N-linked glycans. Contaminants, buffer, salts, and residual peptides were removed from the released glycans by Sep-Pak C18 chromatography and the resulting glycan preparation was permethylated prior to analysis by nanospray ionization mass spectrometry using a linear ion trap (LTQ; Thermo Finnigan). The total ion mapping (TIM) functionality of the Xcalibur software package (version 2.0) was used to obtain total glycan profiles for each sample. Through TIM analyses, automated MS and MS/MS spectra are obtained in small mass increments across a broad range of m/z values. For the analysis of tissue samples, TIM analysis was performed from m/z = 500 - 2000. This mass range collects MS profiles and MS/MS fragmentation spectra for glycans detected from their 1+ to 4+ charge states. Following data collection, resulting TIM

profiles are filtered for the presence of characteristic glycan fragments in the associated MS/MS spectra. By plotting the signal intensity of characteristic fragments as a function of elapsed scan time, a TIM chromatogram is generated that indicates the relative prevalence of specific glycan structures. For assessing the presence of $\beta(1,6)$ -branched glycans extended with at least two N-acetylglucosamine repeats, TIM scans were filtered for the loss of a Hex-HexNAc-Hex-HexNAc fragment from the parent ion.

Proteomic Data Analysis. The raw peptide data was converted to mzXML using ReAdW. MS/MS spectra were searched against the International Protein Index (IPI) human sequence database (IPI.HUMAN.v.3.26, 67 665 sequences) using MyriMatch (Hakomori, 1999. *Biochim. Biophys. Acta* 1473:247–266). The MyriMatch search criteria included only tryptic peptides, all cysteines were presumed carboxyamidomethylated, and methionines were allowed to be oxidized. MyriMatch searches allowed a precursor error of up to 1.25 m/z and a fragment ion limit of within 0.5 m/z. All ambiguous identifications that matched to multiple peptide sequences were excluded. The identified proteins (2+ peptides required) from each individual tumor and normal sample were filtered and grouped using IDPicker software (Zhang et al., 2007. *J Proteome Res.* 6:3549-57). IDPicker software incorporates searches against a reverse database, probability match obtained from MyriMatch, and DeltCN scores to achieve false discovery rates typically < 5%. Information about IDPicker tools can be found at <http://www.mc.vanderbilt.edu/msrc/bioinformatics/>. The raw data files were also analyzed using the TurboSequest algorithm (Peng et al., 2003. *J Proteome Res* 2:43-50; Yates et al., 1995. *Anal Chem* 67:1426-36; Yates et al., 1996. *Analyst* 121:65R-76R) to achieve a false discovery rate of less than 0.3 % for proteins assigned by 2+ peptides using an inverted database (BioWorks 3.1, Thermo Finnigan). Our results indicate that the final MyriMatch/IDPicker proteins list and TurboSequest proteins list showed near complete agreement. All proteins reported in this manuscript were identified using both methods. We found no evidence of amino acid carbamylation following urea elution and DTT reduction.

Biological Function Annotation. Proteins (defined by 2 or more peptides) showing differential binding to L-PHA for tumor compared to normal in at least 2 cases were converted to gene symbols and uploaded to DAVID 2007 (the Database for Annotation, Visualization and Integrated Discovery) for analysis.

Western Blot Experiments. One hundred micrograms of delipidated protein powder were solubilized in 1XTBS/1% triton X-100/ protease inhibitor tablet for

precipitation using antiperiostin (Abcam) (1 μ g). Bound proteins were captured using protein G plus agarose or streptavidin paramagnetic particles before gel electrophoresis and transfer to PVDF membrane prior to probing using biotinylated L-PHA (1:5,000). For L-PHA precipitations 500 μ g of delipidated protein powder was solubilized in 1X TBS/1 % tritonX-100/ protease inhibitors before adding 10 μ g of biotinylated L-PHA and mixing overnight at 4°C. Magnetic streptavidin beads (100 μ l) were used to pull down the L-PHA bound complexes. After washing the beads, proteins were released by boiling in sample loading buffer and separated on 4-12 % NuPage Bis-Tris gels and transferred to PVDF membrane before detection using either anti-periostin Ab (1:1,000) (Abcam) or anti- haptoglobin Ab (1:200) (Santa Cruz Biotechnology). Blots were incubated with anti-rabbit HRP (1:5,000) (Santa Cruz Biotechnology) or anti-mouse HRP (1:5,000) (Santa Cruz Biotechnology) or streptavidin-HRP (1:5,000) (Vector Labs) before washing and detection using Western Lightening Plus (Perkin Elmer).

Results

An initial set of 5 human ductal invasive breast carcinoma tissue samples were used to evaluate the use of the lectin L-PHA to bind and enrich for potential glycoprotein biomarkers to distinguish breast cancer from normal tissue. As shown in Table 3 these cases represent metastatic and non-metastatic disease, cases positive and negative for amplified her2/neu, and cases that were both estrogen receptor (ER) and progesterone receptor (PR) positive and negative. Our approach is to analyze breast tissue, comparing normal and tumor tissue from the same patient, to identify potential glycoproteins that react with the lectin L-PHA. All cases showed increased levels of L-PHA binding indicative of β (1,6) branched N-linked glycans except case 10406, therefore, case 10406 was not analyzed by NSI-MS. We conclude that 4 out of 5 cases or 80 % of the tumor tissue analyzed had proteins with increased levels of β (1,6) branched N-linked glycan structures relative to normal breast tissue. To isolate these L-PHA reactive glycoproteins, we developed a method using intact proteins for the lectin binding, which differs from the more common method of using glycopeptides (Fig. 3). In addition, we found that delipidation of the breast tissue prior to analysis significantly improved MS/MS results (Fig. 3) (Aoki et al., 2007. *J Biol Chem* 282:9127-42).

Table 3. Summary of cases analyzed

Case	Unique Proteins Total NL	Unique Proteins Total TU	Unique Proteins L-PHA-NL	Unique Proteins L-PHA-TU	Tumor Grade	HER2 Status	ER Status	PR Status	LN Status	L-PHA Level
10406	ND	ND	ND	ND	II, moderate	Positive	3+ Positive	Negative	Yes	1
10119	362	363	88	80	II, moderate	Negative	3+ Positive	3+ Positive	No	4
11827	349	515	70	118	III, poor	ND	ND	ND	Yes	4
2417	347	491	53	214	III, poor	2+ Positive	Negative	Negative	Yes	10
2207	362	476	145	193	III, poor	2+ Positive	3+ Positive	3+ Positive	No	9

To verify that L-PHA binding accurately reports intrinsic differences in glycan expression between tumor and adjacent normal tissue, total N-linked glycans were profiled for tissues taken from case 2417, which exhibited the greatest relative increase in tumor-associated L-PHA recognition (Table 3). The prevalence of glycans carrying a $\beta(1,6)$ branch extended with N-acetylactosamine was compared in tumor and adjacent normal preparations by quantifying the signal intensity of a specific fragment ion detected in TIM analysis. By filtering TIM profiles for the presence of a permethylated Hex-HexNAc-Hex-HexNAc fragment, the relative abundances of several $\beta(1,6)$ branched parent ions were measured and compared. The most prevalent of the detected glycans extended with N-acetylactosamine was increased more than 2.5-fold in tumor tissue relative to adjacent normal tissue, compared using equivalent protein amounts as starting material (Fig. 4). The trimmed, high mannose glycan Man5GlcNAc2, that serves as an early precursor for complex terminal modifications including $\beta(1,6)$ branching, was found in equal prevalence between the two tissues (data not shown). Therefore for case 2417, the increased abundance of L-PHA reactive proteins identified by MS/MS from tumor tissue compared with normal tissue correlates with increases in $\beta(1,6)$ branched N-glycan structures, confirming the specificity of the lectin enrichment analysis.

Proteins Showing Tumor-Specific Increased Reactivity with L-PHA. We detected a total of 258 L-PHA reactive proteins (2 or more peptides) from the 4 tumors showing elevated L-PHA reactivity that were analyzed by MS/MS. For each of the 258 unique proteins identified, we examined the number of peptides identified for that protein from NSI-MS/MS analysis of normal tissue and tumor tissue before and after L-PHA fractionation. Differences in protein abundance between normal and tumor were normalized by determining the ratio of peptides identified in patient matched normal and tumor tissue prior to L-PHA fractionation. Proteins were eliminated in each case if they did not show a minimum of 1.5 fold increase in peptides identified from tumor relative to normal after L-PHA fractionation. From the list of proteins showing at least a 1.5 fold

increase in tumor relative to normal, proteins were considered “enriched” following LPHA fractionation if they were identified from at least 2 cases. A total of 34 proteins had increased peptides and spectra present in L-PHA fractionations isolated from tumor compared with normal tissue for at least 2 separate cases of ductal invasive breast carcinoma (Table 4, number of spectra in parenthesis). The peptide sequences for these proteins are provided in Table 4. As expected, the majority of these proteins are predicted to be glycoproteins by searching databases such as GenBank and IPI. Since we do not compete L-PHA bound proteins from the magnetic beads using a competitive sugar hapten, we have identified some proteins that are predicted to be non-glycosylated and are likely binding to L-PHA reactive glycosylated proteins. We utilized DAVID 2007 (Database for Annotation, Visualization and Integrated Discovery) to annotate the function of the proteins listed in Table 4 (Fig. 5A). The top 2 functional classifications of glycoproteins showing differential binding to L-PHA in ductal invasive breast carcinoma relative to normal breast tissue are (i) binding proteins related to cell adhesion and cell-communication, and (ii) proteins that respond to external stimulus. These results indicate that largest group of proteins showing increases in $\beta(1,6)$ branched glycosylation function in mediating communication of the tumor cell with the extracellular matrix (COL6A1, COL6A2, COL6A3, COL14A1) and neighboring cells (HPX, BGN, DCN, VIM, POSTN, VTN, THBS1, OGN). The second highest group of proteins with elevated L-PHA reactive glycoproteins participates in the response of the tumor cells to environmental stress. This group includes the members of the lectin-induced complement pathway (C3 and C4B), activators of complement (HPR), immunoglobulin/MHC complex (IGHA1, IGLV4-3, IGHM), enzyme inhibitors (SERPINA1, 14-3-3 zeta/delta, COL6A3, THBS1), enzyme activators of MMP (HPX), and enzymes that detoxify (PRDX1). This information contributes to our knowledge of the functions of glycoproteins acquiring $\beta(1,6)$ branched N-linked glycan structures in breast carcinoma and offers insights into how the acquisition of these structures may be associated with breast cancer progression. Several of the proteins enriched by L-PHA are predicted to participate in the development of the skeletal system and organ development (POSTN, ANXA2, DCN, COL6A3, PRDX1, and OGN). There are also several enzymes that participate in the glycolytic pathway that are enriched by L-PHA (PGK1, TPI1, ENO1, LDHA, and PKM2). Overall, these results suggest that proteins that are participating in cell communication, organ development, and metabolism derive $\beta(1,6)$ branched N-linked glycan structures in breast tumors. The abnormal elevated expression of these glycan structures on these proteins in

breast epithelial cells may play roles in tumor progression and invasion (Guo et al., 2007.

J Biol Chem 282:22150-62).

Table 4. L-PHA reactive proteins with elevated peptides and spectra in tumor relative to normal for at least 2 cases.

Abb name	Name	10119-NL	10119-TU	11827-NL	11827-TU	2207-NL	2207-TU	2417-NL	2417-TU	Function	Cell Compartment
PRDX1	Peroxisiredoxin-1	0	0	0	1(1)	1(2)	2(3)	0	0	proliferation	Unknown
POSTN	Periostin	0	4(9)	0	6(14)	0	11(26)	1(1)	10(51)	adhesion	Ext
BGN	Biglycan	4(7)	1(3)	0	0	3(7)	6(12)	1(3)	7(14)	binding	Ext
CLIC1	Chloride intracellular channel protein	0	0	0	0	0	1(2)	0	1(2)	Cl transport	PM
DCN	Decorin	3(7)	1(5)	0	0	4(14)	6(18)	2(6)	8(18)	organ morpho	Ext
KCIP-1	14-3-3 zeta/delta	0	1(1)	0	3(5)	0	3(6)	0	0	binding	Cyto
APOA-1	Apo-A1 precursor	1(1)	2(4)	3(7)	2(4)	4(7)	1(1)	1(1)	5(8)	binding	Ext
HPX	Hemopexin	0	0	0	0	3	5	0	6	binding	Ext
OGN	Osteoglycin	3(7)	5(8)	0	1(2)	0	8(25)	0	7(12)	proliferation	Unknown
COL6A3	Collagen VI alpha 3	22(56)	25(47)	10(14)	7(10)	40(105)	60(167)	26(167)	86(290)	adhesion	Ext
C3	187 kD protein	4(8)	1(2)	0	0	4(6)	11(18)	0	25(44)	binding	Ext
PGK1	Phosphoglycerate kinase 1	0	0	0	1(2)	0	2(3)	0	3	catalytic	Cyto
COL14A1	Collagen 14 isoform 1	1(2)	9(15)	0	0	14(43)	22(72)	6(13)	28(98)	cell-cell adhesion	Ext
PROF1	Profilin	0	0	0	1(2)	0	1(4)	0	2(3)	binding	Cyto
LDHA	Lactate Dehydrogenase isoform 1	0	0	0	1(2)	0	3(6)	0	1(1)	catalytic	Ext
TUBA1C	Tubulin alpha 6	0	1(1)	0	2(5)	3(3)	7(18)	1(1)	7(12)	binding	Cyto
COL6A1	Collagen alpha 1 VI chain precursor	6(13)	5(0)	1(2)	2(3)	7(23)	13(36)	2(3)	14(43)	adhesion	Ext
THBS1	Thrombospondin-1	0	0	0	3(4)	0	0	0	1(2)	motility	Ext
VTN	Vitronectin	0	1(1)	0	1(1)	0	0	0	2(3)	adhesion	Ext
COL6A2	Collagen alpha 1 VI isoform 2C2	7(19)	8(15)	5(7)	3(7)	10(22)	12(31)	5(13)	14(62)	cell-cell adhesion	Ext
HSPA1	Heat Shock 70 kDa	0	0	0	1(1)	0	1(1)	0	0	unfolded protein	Ext
ANXA5	Annesin A5	0	0	0	2(2)	0	1(2)	0	0	binding	Cyto
IGLV4-3	variable Ig	0	1(1)	0	1(1)	3(5)	5(12)	0	8(32)	immune res	Ext
C4B	C4B complement	0	0	0	0	1(2)	3(7)	0	3(5)	binding	Ext
PPIA	Peptidyl prolyl isomerase A	0	0	0	2(5)	0	0	0	1(2)	catalytic	
IGHA1	MHC Class 1 protein	0	0	0	0	2(5)	4(10)	2(2)	7(14)	binding	Ext
ANXA2	Annexin A2	0	0	1(1)	2(2)	2(3)	5(14)	0	0	enzyme inhibitor	PM
TPI1	Triphosphate iso	0	1(1)	0	1(2)	4(3)	3(4)	0	3(3)	carb metabolism	Cyto
ENO1	Alpha enolase	0	0	0	3(5)	4(17)	3(5)	0	5(10)	glycolytic	Cyto
IGHM	IGHM protein	0	0	0	0	8(37)	10(54)	6(13)	11(63)	immune res	PM
HPR	Haptoglobin-related protein precursor Isoform-1	0	1(1)	0	1(2)	0	1(2)	0	3(4)	defense	Ext
PKM2	Pyruvate kinase M2	0	0	0	0	1(1)	5(12)	0	3(9)	binding	Cyto
SERPINA1	Alpha-1 antitrypsin inhibitor	0	1(1)	2(4)	2(4)	4(11)	4(12)	1(1)	8(17)	binding	Ext
VM	Vimentin like 50 kDa	1(1)	3(6)	6(9)	6(12)	2(3)	5(8)	0	5(6)	motility	Cyto

^a International protein index database # of spectra in parenthesis beside the number of peptides

^b Gene function determined by the Gene Ontology Consortium

Table 5. Proteins with increased peptides and spectra present in L-PHA fractions isolated from tumors.

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
1	IPI00000874	PRDX1	22.1	2	9.5	ADEGISFR	1	1
	IPI00000874	PRDX1	22.1	2	9.5	QITVNDLPVGR	2	3
2	IPI00022488	HPX	51.6	7	22	NFPSPVDAAFR	3	1
						LLQDEFPGIPSPLDAAVECHR	4	1
						YYCFQGNQFLR	5	2
						GGYTLVSGYPK	6	1
						EVGTPHGIILDSVDAAFICPGSSR	7	1
						RLWWDLK	8	1
						SGAQATWTELPWPHEK	9	4
3	IPI00007960	POSTN	93.3	11	21.8	DQGPNCALQQILGTK	10	2
						LREEIEGK	11	1
						VLTIQTSIQDFIEAEDDLSSFR	12	18
						AAAITSDILEALGR	13	7
						DGHFTLFAPTNEAFK	14	9
						DIVTNGVIHLIDQVLIPDSAK	15	5
						VGLNELYNGQILETIGGK	16	8
						FSTFLSLLEAADLK	17	7
						ELLTQPGDWTLFVPTNDAFK	18	10
						LLYPADTPVGNDQLLEILNK	19	11
4	IPI00216691	PROF1	15	2	20	IIDGVPVEITEK	20	2
						DSPSVWAAVPGK	21	1
5	IPI00072917	COL6A3	322	77	36.2	TFVNITPAEVGVLVGK	22	8
						VGLVQFSDTPVTEFSLNTYQTK	23	2
						SDILGHLR	24	2
						TLSGTPEESKR	25	2
						AAPLQGLPGLLAPLR	26	4
						LLPYIVGVAQR	27	2
						MKPLDGSALYTGSAIDFVR	28	1
						NNLFTSSAGYR	29	8
						LLVLITGGK	30	2
						SLDEISQPAQELK	31	6
						GADQAELEEIAFDSSLVFIPAEFR	32	3
						DILFLFDGSANLVGQFPVVR	33	11
						IIDELNVKPEGTR	34	4
						IAVAQYSDDVKVESR	35	3
						FDEHQSKPEILNLVK	36	5
						ALNLGYALDYAQR	37	16
						SSDRVDGPASNLK	38	5
						VDGPASNLK	39	1
						QSGVVPFIFQAK	40	3
						NADPAELEQIVLSPAFILAAESLPK	41	14
						IGDLHPQIVNLLK	42	4
						DVVFLLDGSEGVV	43	4
						SGFPLLK	44	4
						VVESLDVGQDR	45	9
						VAVVQYSR	46	4
						QLTLGGPTPNTGAALFVLR	47	9
						NILVSSAGSR	48	11
						ITEGVPQLLIVLTADR	49	9
						SGDDVRNPSVVVK	50	6

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
						QLGTVQQVISER	51	9
						VTQLTR	52	1
						LQPVLQPLPSGVGGK	53	9
						DVVFLIDGSQSAGPEFQYVR	54	15
						LVDYLDVGFDTTR	55	29
						VAVIQFSDDPK	56	4
						VEFLNAHSSKDEVQNAVQR	57	1
						DEVQNAVQR	58	4
						QINVGNALYVSR	59	12
						IEEGVPQFLVLISGK	60	8
						SDDEVDDPAVELK	61	8
						QFGVAPFTIAR	62	3
						NADQEELVK	63	14
						ISLSPEYVFVSTFR	64	7
						LLTPITLTSEQIQK	65	14
						RLNIGPSK	66	8
						LNIGPSK	67	1
						VGVVQFSNDVFPEFYLK	68	7
						SQAPVLDAIR	69	11
						ALEFVAR	70	2
						IEDGVPQHLVLVGK	71	5
						SSGIVSLGVGDR	72	5
						NIDRTELQTITNDPR	73	2
						TELQTITNDPR	74	7
						LVFTVR	75	2
						DSFQEVLR	76	9
						RQIIDAINK	77	1
						VGLEHLR	78	3
						VNHFVPEAGSR	79	1
						VPQIAFVITGGK	80	10
						SVEDAQDVSLALTQR	81	9
						VFAVGVR	82	10
						NIDSEEVGK	83	10
						IASNSATAFR	84	5
						ACNLDVILSFDGSR	85	6
						DQNVFVAQK	86	8
						VSVVANTPSGPVEAFDFEYQPEMLEK	87	3
						SQHPYVLTEDTLK	88	5
						VVIHFTDGADGLADLHR	89	5
						ALILVGLER	90	9
						VVNLER	91	7
						LNLLDLDYELAEQLDNIAEK	92	18
						GETGDDGRDGVGSEGR	93	5
						GDSIDQCALIQSIK	94	1
						DVLSIVNDLTIAESNCPR	95	6
						VAVVTYNNEVTTEIR	96	14
						NLQVALTSK	97	4
						VAVFFSNTPTR	98	2
						ALGSAIEYTIENVFESAPNPR	99	7
						LLDSFVSSENAFYLSPIR	100	6
6	IPI00218343	TUBA1C	49.8	9	30.5	LIGQIVSSITASLR	101	2
						TIQFVDWCPTGFK	102	1

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
						LISQIVSSITASLR	103	6
						TIGGGDDSFNTFFSETGAGK	104	2
						AVFVDLEPTVIDEVR	105	10
						EIIDLVLDR	106	4
						NLDIERPTYTNLNR	107	1
						FDGALNVDLTEFQTNLVPYPR	108	8
						VGINYQPPTVVPGGDLAK	109	4
7	IPI00025465	OGN	33.9	7	29.2	DFADIPNLR	110	9
						RLDFTGNLIEDIEDGTFSK	111	9
						LSLLEELSLAENQLLK	112	11
						LPVLPPK	113	1
						LIHLQFNNIASITDDTFCK	114	2
						DRIIEIR	115	2
						LEGNPIVLGK	116	10
8	IPI00477597	HPR	39	3	8.6	DIAPTLTYVGK	117	1
						NPANPVQR	118	1
						VTSIQDWVQK	119	7
9	IPI00479186	PKM2	57.9	6	18.6	LDIDSPITAR	120	2
						NTGICTIGPASR	121	2
						TATESFASDPILYRPVAVALDTK	122	2
						LAPITSDPTEATAVGAVEASFK	123	10
						APIIAVTR	124	3
						GIFPVLCKDPVQEAEDVDLR	125	2
10	IPI00465248	ENO1	47.1	4	20	AAVPSGASTGIYEALER	126	6
						DATNVGDEGGFAPNILENKEGLELLK	127	1
						YISPDQLADLYK	128	6
						SGETEDTFIADLVVGLCTGQIK	129	2
11	IPI00164623	C3	187.3	24	19.9	TIYTPGSTVLYR	130	2
						AYYENSPOQVFSTFEVK	131	6
						EYVLPSEFIVEPTEK	132	5
						IPIEDGSGEVLSR	133	4
						VLLDGVQNP	134	1
						VPVAVQGEDTVQSLTQGDGVAK	135	2
						TKKQELSEAEQATR	136	1
						TELRPGETLNVNLLR	137	1
						EPGQDLVVLPLSITDFIPSFR	138	2
						LVAYYTLIGASGQR	139	2
						IWDVVEK	140	2
						ASHLGLAR	141	1
						SNLDEDIAEENIVSR	142	5
						LPYSVVR	143	2
						NEQVEIR	144	6
						SSLSVPYVIVPLK	145	3
						TGLQEEVK	146	6
						ILLDGTTPVAQMTEDAVDAER	147	1
						DFDFVPPVVR	148	3
						VHQYFNVELIQGAVK	149	2
						ACEPGVDYVYK	150	1
						VQLSNDFDEYIMAEQTIK	151	1
						SGSDEVQVGGQR	152	3
						DTWVEHWPEDECQDEENQK	153	2
12	IPI00021841	APOA1	30.7	4	18.9	DYVSQFEGSALGK	154	4

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
						LLDNWDSVTSTFSK	155	4
						QGLLPVLESFK	156	5
						VSFLSALEEYTK	157	4
13	IPI00021263	KCIP-1	27.7	4	23	SVTEQGAELSNEER	158	5
						DICNDVLSLLEK	159	2
						YLAEVAAGDDKK	160	1
						TAFDEAIAELDTLSEESYK	161	4
14	IPI00010790	BGN	41.6	8	26.9	VVQCSDLGLK	162	3
						EISPDTTLLDLQNNDISLR	163	8
						NHLVEIPPNLPSLVELR	164	7
						GVFSGLR	165	1
						DLPETLNELHLDHMK	166	1
						IQAIELEDLLR	167	9
						LGLGHNQIR	168	2
						VPSGLPDLK	169	3
15	IPI00012119	DCN	39.7	8	29.8	DFEPSLGPVCPFR	170	4
						VVQCSDLGLDKVPK	171	1
						DLPPTTLLDLQNNK	172	23
						ITEIKDGDFFK	173	1
						NLHALILVNNK	174	4
						VSPGAFTPLVK	175	2
						VPGLAEHK	176	3
						ASYSGVSLFSNPVQYWEIQPSTFR	177	1
16	IPI00382938	IGLV4	25.9	5	33.3	AAPSVTLFPPSSEELQANK	178	16
						AGVETTTPSK	179	9
						ATLVCLISDFYPGAVTVAWK	180	3
						YAASSYLSLTPEQWK	181	5
						SYSCQVTHEGSTVEK	182	4
17	IPI00418163	C4B	192.7	5	4.2	LNMGITDLQGLR	183	1
						VGDTLNLNLR	184	2
						SFFPENWLWR	185	1
						VTASDPLDTLGSEGALSPGGVASLLR	186	5
						VLSLAQEQVGGSPK	187	3
18	IPI00430842	IGHA1	52.8	7	16.8	DASGVFTFTWPSSGK	188	4
						SAVQGPPER	189	2
						TFTCTAAYPEK	190	2
						TPLTATLSK	191	5
						WLQGSQELPR	192	5
						YLTWASR	193	2
						QEPSQGTTFVAVTSILR	194	2
19	IPI00465028	TPI1	30.7	2	9.4	VVLAYEPVWAIGTGK	195	5
						SNVSDAVAQSTR	196	5
20	IPI00472610	IGHM	52.6	14	45	GPSVFPLAPSSK	197	6
						GTTVTVSSASTK	198	5
						THTCPPCPAPELLGGPSVFLFPPKPK	199	1
						TPEVTCVVVDVSHEDPEVK	200	12
						FNWYVDGVEVHNAK	201	9
						TTPPVLDSDGSFFLYSK	202	23
						STSESTAALGCLVK	203	2
						GFYPDSIAVEWESNGQPENNYK	204	17
						TPEVTCVVVDVSHEDPEVQFK	205	1

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
						WYVDGVEVHNAK	206	2
						VVSVLTVLHQDWLNGK	207	2
						DTLMISR	208	10
						GPSVFPLAPCSR	209	1
						NQCSLTCLVK	210	4
21	IPI00827679	VIM	50	10	27.8	SLYASSPGGVYATR	211	1
						LLQDSVDFSLADAINTEFK	212	6
						VELQELNDR	213	1
						ILLAELEQLK	214	6
						EEAENTLQSFR	215	1
						QDVDNASLAR	216	1
						NLQEAEEWYK	217	6
						FADLSEAANR	218	5
						QVQSLTCEVDALK	219	2
						ISLPLPNFSSLNLR	220	3
22	IPI00169383	PGK1	44.6	3	11	YSLEPVAVELK	221	1
						ACANPAAGSVILLENLR	222	2
						QIVWNGPVGVFWEAFAR	223	2
23	IPI00176193	COL14A1	193.5	26	18.9	TNQLNLQNTATK	224	8
						HFLNLVTAFDVGSEK	225	6
						DEVIEAVR	226	10
						IGILITDGK	227	4
						SQDDIIPPSR	228	4
						ASAHITGPTELITSEVTAR	229	11
						WDAVTGASGYLILYAPLTEGLAGDEK	230	1
						ISNVGSNSAR	231	4
						IVYNNADGTEINEVEVDPIITFPLK	232	8
						NLVVGDETTSSLR	233	6
						WDISDSOVQQFR	234	5
						VTVTPIYTDGEGVSVSAPGK	235	14
						TLPSSGPQNLR	236	7
						VSEEWYNR	237	2
						ITWDPSPSPVK	238	4
						TLFLGVTNLQAK	239	7
						VVIESLQDR	240	5
						IISFLYSTVGALNK	241	4
						TKETLLDAIK	242	4
						DTLFTAESGTR	243	6
						VIVVITDGR	244	4
						HVFFVDDFDAFK	245	2
						DGIDLAGFK	246	4
						ILPDTQPFPALWEILNK	247	11
						NSDPLVGILDNGGK	248	13
24	IPI00217966	LDHA	36.6	2	8.1	DLADELALVDVIEDK	249	4
						VIGSGCNLDSAR	250	4
25	IPI00291136	COL6A1	108.5	17	24.8	DAEEAISQTIDTIVDIK	251	1
						RFIDNLR	252	2
						FIDNLR	253	1
						YLIVVTDGHPLEGYKEPCGGLEDVNEAK	254	3
						VFSVAITPDHLEPR	255	4
						LSIATDHTYR	256	3
						SGDEGPPGSEGAR	257	6

Number	IPI	Name	M Wt (kDa)	# Unique Peptides	% Coverage	Peptide Sequence	SEQ ID NO:	Spectra
						SLQWMAGGTFTEALQYTR	258	3
						IALVITDGR	259	9
						DTTPLNVLCSPGIQVSVGIK	260	5
						DVFDIPGSDQLNVISCOGLAPSQGRPGLSLVK	261	3
						ENYAELLEDAFLK	262	8
						LLLFSDGNSQGATPAAIEK	263	21
						AGIEIFVVVGR	264	1
						TAEYDVAYGESHLFR	265	6
						VPSYQALLR	266	8
						GVFHQTVSR	267	2
26	IPI00296099	THBS1	129.3	4	5.1	IEDANLIPPVPDDKFQDLVDAVR	268	2
						FVFGTTPEDILR	269	1
						TIVTTLQDSIR	270	2
						KVTEENKELANELR	271	1
27	IPI00298971	VTN	54.3	2	5.6	DVWGIEGPIDAAFR	272	4
						FEDGVLDPDYPR	273	1
28	IPI00304840	COL6A2	108.5	13	15.2	NLQGISSFR	274	10
						LFAVAPNQNLK	275	12
						DIASTPHELYR	276	7
						NDYATMLPDSTEIDQDTINR	277	3
						NFVINVVNR	278	6
						NLEWIAGGTWTPSALK	279	14
						VFAVVITDGR	280	7
						DDDLNLR	281	9
						HESENLYSIACDKPQQVR	282	1
						LGEQNFHK	283	9
						FVEQVAR	284	6
						RDDDPLNAR	285	12
						AAVFHEKDYDSLAPGFFDR	286	2
29	IPI00304925	HSPA1	70	2	6.8	DAGVIAGLNVL	287	1
						ELEQVCNPIISGLYQGAGGPGGFGAQQGPK	288	1
30	IPI00419585	PPIA	18	2	10.9	VSFELFADK	289	3
						FEDENFILK	290	4
31	IPI00553177	SERPINA1	46.7	8	24.8	SVLGQLGITK	291	3
						VFSNGADLSGVTEEAPLK	292	10
						ITPNLAEFASFSLYR	293	1
						AVLTIDEK	294	1
						VVNPTQK	295	1
						TLNQPDSQLQLTTGNGLFLSEGLK	296	11
						LQHLENELTHDIITK	297	5
						SASLHLPK	298	2
32	IPI00010896	CLIC1	26.9	2	14.5	LAALNPESNTAGLDIFAK	299	2
						FLDGNELTLADCNLLPK	300	2
33	IPI00455315	ANXA2	38.6	5	20	GVDEVTIVNLTNR	301	4
						TPAQYDASELK	302	5
						GLGTDEDSLIEIICSR	303	2
						TNQELQEINR	304	4
						TDLEKDIISDTSGDFRK	305	1
34	IPI00329801	ANXA5	35.9	2	8.7	GLGTDEESILTLLTSR	306	1
						DLLDDLKSELTKGK	307	2

Interestingly, many of the proteins identified in Table 4 are connected with TGF β signaling pathways. A recent manuscript identifies increased TGF β protein levels in breast tumor tissue as a factor that correlates with shorter disease-free survival (Desruisseau et al., 2006. *Br J Cancer* 94:239-46). Many of the proteins we identified after L-PHA enrichment in tumor tissue are either induced by TGF β (POSTN, COL6A3, SERPINA1) or are known to bind TGF β with nanomolar affinities (BGN, DCN). It will be interesting to determine how β (1,6) branched N-linked glycosylation influences the binding of TGF β to these proteins.

L-PHA capture facilitates the identification of biomarkers likely to be found in serum. We have examined the predicted cellular distribution of proteins enriched by L-PHA listed in Table 4. The distribution is primarily extracellular (56 %) and cytoplasmic (29 %) with the remainder of proteins localized on the plasma membrane (9 %) or unknown (6 %) (Fig. 5B). L-PHA-affinity enrichment significantly concentrates proteins identified from the extracellular region. For comparison, only 14-15 % of the total proteins isolated from normal and tumor before L-PHA enrichment are extracellular (Fig. 5C and 5D).

L-PHA enrichment can identify novel markers for breast carcinoma.

Approximately 73 % of the 34 tumor L-PHA enriched proteins have been previously reported in breast cancer studies (n=25); while 29 % (n=9) have not been previously cited for breast cancer (Table 4). One of the novel, tumor-specific L-PHA reactive glycoproteins that we identified, known as osteoglycin (OGN), was present at similar levels of peptide abundance in normal and tumor breast tissue before L-PHA fractionation. However, OGN is consistently identified from tumor tissue after L-PHA fractionation. Osteoglycin has not been identified as a potential marker for breast cancer previously, likely due to the consistent levels of protein present in normal and tumor. Another interesting protein not previously associated with breast cancer is the 14-3-3 zeta protein. This protein is not predicted to be glycosylated and is probably enriched by L-PHA due to association with a protein that binds L-PHA. 14-3-3 zeta functions as an adapter protein that binds with other proteins controlling cell growth and proliferation. The cellular localization of 14-3-3 zeta protein has been reported to be largely cytoplasmic, but it has been reported to be present on the plasma membrane and in the Golgi (Fu et al., 2000. *Annu Rev Pharmacol Toxicol* 40:617-47). We find in 3 out of the 4 cases significant tumor-specific association with L-PHA for 14-3-3 zeta. Similar levels of 14-3-3 zeta present in normal and tumor tissue probably prevented previous identification

of this protein as a marker for breast cancer. We conclude that selective enrichment using the lectin L-PHA has enabled the identification of novel markers for breast carcinoma and adds an additional level of biomarker selection.

L-PHA enrichment increases the identification of markers common to breast carcinoma cases with diverse clinical features. The quest to identify markers for the early detection of many tumors has been hampered by tumor heterogeneity. Our approach, focusing on a specific post-translational modification that increases in parallel with malignant progression, has enabled the identification of 12 markers common to all 4 cases of breast carcinoma analyzed (Fig. 6). A key element of our success is the targeting of $\beta(1,6)$ branched N-linked glycan structures that are normally expressed at a low level in non-diseased breast epithelial cells. This targeted approach has enabled the identification of proteins that change glycosylation only in breast carcinoma tumor tissue.

Validation of glycoproteins with differential L-PHA reactivity in normal and tumor tissue. We have selected 2 markers for further validation from Table 4. These glycoproteins were selected due to enrichment in tumor tissue for all cases analyzed in Table 4.

Periostin. Periostin (POSTN) has been identified as a possible factor promoting breast cancer progression through induction of angiogenesis (Shao et al., 2004. *Mol Cell Biol* 24:3992-4003). Previous studies analyzing mRNA levels by pooled RNA sampling and immunohistochemistry arrays indicated that POSTN mRNA and protein levels were increased in malignant breast epithelial cells (Shao et al., 2004. *Mol Cell Biol* 24:3992-4003; Grigoriadis et al., 2006. *Breast Cancer Res* 8:R56). We found in the total MS/MS analysis of matched normal and malignant breast tissue before L-PHA fractionation that there were roughly twice as many POSTN peptides identified from tumor tissue compared with normal tissue for cases 10119 and 2207. For case 11827, we did not detect POSTN before L-PHA fractionation and for 2417 we found equivalent levels of POSTN peptides identified in normal and tumor tissue before L-PHA enrichment using NSI-MS/MS (Fig. 7A). These results are not quantitative and are only a qualitative assessment of abundance. However, they suggest that examining alterations in the abundance of POSTN protein alone would not be a selective marker for breast carcinoma. However, we identified POSTN peptides after L-PHA enrichment and MS/MS analysis in all 4 of the tumor tissues analyzed with only 1 peptide identified in normal tissue for case 2417 (Fig. 7A). The high degree of tumor-specific association of POSTN with LPHA suggests that the presence of $\beta(1,6)$ branched N-linked glycosylation on POSTN is a marker of breast

cancer. To confirm these results, we have immunoprecipitated POSTN from normal and tumor breast tissue using an anti-periostin antibody and probed the blot using biotinylated LPHA. Tissue amounts were limiting for case 10119 preventing further analysis of this case. For each case analyzed, POSTN reactivity with L-PHA is higher compared with the matched normal tissue control (Fig. 7B, panel 1). To normalize for total periostin protein, we probed the membrane using the anti-periostin antibody (Fig. 7B panel 2). To confirm these results, we precipitated using biotinylated L-PHA followed by detection using the anti-periostin antibody (Fig. 7B panel 3). To quantitate the relative increase in L-PHA reactive periostin, films for total periostin and LPHA reactive periostin were scanned and analyzed by densitometry (Fig. 7C). Overall, our results indicate that POSTN acquires increased levels of $\beta(1,6)$ branched N-linked structures in the breast tumor tissue relative to normal breast tissue.

Haptoglobin-related protein precursor. Haptoglobin-related precursor protein (HPR) has 90 % sequence identity to the conventional form of haptoglobin (HP) found in serum. The adult liver expresses low levels of HPR mRNA (Bensi et al., 1985. *EMBO J* 4:119-26) and very little HPR protein has been detected in serum (Fawcett et al., 1990. *Biochim Biophys Acta* 1048:187-93), suggesting that the HPR we detected in breast carcinoma tissue is likely expressed by breast cancer epithelial cells. HPR shares antigenic epitopes with the pregnancy-associated plasma protein-A that is secreted from uterine epithelial cells or placenta during pregnancy and has been reported as an independent prognostic factor useful for detecting the recurrence of breast cancer (Kuhajda et al., 1989. *Proc Natl Acad Sci U S A* 86:1188-92; Kuhajda et al., 1989. *N Engl J Med* 321:636-41). To validate the identification of L-PHA reactive HPR as a breast tumor-specific marker we performed L-PHA precipitation followed by Western blotting. Due to the fact that we did not detect HPR prior to L-PHA affinity enrichment (Fig. 7D) we have chosen to validate using L-PHA precipitation followed by detection using an anti-haptoglobin antibody that recognizes HP as well as HPR (Fig. 7E). Our results revealed an increased association of the beta chain of haptoglobin with L-PHA for the tumor tissue compared with normal breast tissue in cases 2417 and 2207. For case 11827, we observed a similar intensity in beta haptoglobin between normal and tumor on the Western blot. In all 3 cases, the tumor haptoglobin displayed a shift to a larger molecular weight compared with normal breast tissue from the same patient. Therefore, we conclude that in agreement with our MS/MS data, tumor haptoglobin has elevated $\beta(1,6)$ branched N-linked glycan structures compared with normal.

Discussion

Elevation of $\beta(1,6)$ branched N-linked glycans in breast cancer has been previously cited as a poor prognostic indicator (Handerson et al., 2005. *Clin Cancer Res* 11:2969-73). In this study we have used L-PHA, a lectin that specifically recognizes these glycan structures, to pull out potential biomarkers for breast carcinoma. Using this type of targeted glycoproteomic approach enabled us to identify markers common to breast cancer tissues with different stages, hormone receptor status, lymph node status, and her2/neu status. Our ability to analyze the relative abundance of biomarkers in normal and tumor tissue from the same patient, before and after L-PHA fractionation, eliminates possible bias that may be introduced from differences in individual gene expression profiles. Our targeted glycoproteomic approach has enabled the identification of several potential markers for breast carcinoma. Therefore, future studies focused on defining the normal and tumor glycome of various tissues can be useful for the development of new lectin targeting strategies to identify glycoproteins with tumor-specific glycan alterations.

TGF β Connection. We have identified several proteins either induced by TGF β , or known to associate with TGF β , suggesting a link between $\beta(1,6)$ branched N-linked glycosylation in breast tumors and the TGF β pathway. Changes in downstream signaling controlled by TGF β have been documented for breast cancer (Gomis et al., 2006. *Cancer Cell* 10:203-14). However, unlike other types of malignancy that have evaded the normal growth-inhibitory functions of TGF β through inactivating mutations in TGF β receptors, the mechanisms of breast cancer resistance to TGF β -mediated growth inhibition remain poorly understood (Massague and Gomis, 2006. *FEBS Lett* 580:2811-20). We have found increased $\beta(1,6)$ branched N-linked glycosylation on several extracellular proteins known to interact with TGF β . Differential glycosylation of these ligands may initiate alterations in TGF β signaling. The extracellular proteoglycans decorin and biglycan have been shown to bind TGF β reducing its bio-availability for TGF β RI and TGF β RII during skeletal muscle differentiation (Droguett et al., 2006. *Matrix Biol* 25:332-41). Therefore, it may be possible that the aberrant glycosylation of small leucine-rich proteoglycans such as decorin (DCN) and biglycan (BGN) increases the amount of TGF β sequestered in the extracellular matrix complexes making less TGF β available for canonical TGF β receptor activation. Increased expression of TGF β in fibroblasts has been shown to induce the expression of several glycoproteins implicated in the pathogenesis of breast cancer such as collagen VI α 3, tenascin, and PAI-1 (Berking et al., 2001. *Cancer Res* 61:8306-16). In

this study, we have identified collagen VI $\alpha 3$ (COL6A3) as one of the proteins enriched by L-PHA in tumor tissue. Collagen VI upregulation and secretion is associated with increased cell survival via resistance to apoptosis through down-regulation of Bax and prevention of $\beta 1$ integrin-mediated apoptosis (Ruhl et al., 1999. *J Biol Chem* 274:34361-8). Logically, increased levels of TGF β in the tumor should initiate the cytostatic effect often associated with TGF β signaling. However, elevated TGF β levels in the tumor can have an opposite effect by inducing collagen VI expression that can oppose cytostatic effects by blocking apoptosis (Ruhl et al., 1999. *J Biol Chem* 274:34361-8). Interestingly, collagen VI has been shown to associate in a ternary complex with DCN, and BGN by co-immunoprecipitation experiments (Reinboth et al., 2006. *J Biol Chem* 281:7816-24). COL6A3, DCN, and BGN were enriched by L-PHA in at least 2 cases (Table 4) suggesting a possible link between the formation of this complex and the presence of $\beta(1,6)$ branched N-linked glycans. Experiments investigating the impact of $\beta(1,6)$ branched glycosylation on the formation of this complex and the resulting effect on cell survival in breast epithelial cells will be necessary to evaluate this hypothesis as a possible mechanism for breast tumor cells evasion of TGF β -induced cytostatic response.

An enrichment of L-PHA reactive glycoproteins that have functions in the skeletal system. Many of the proteins that were enriched by L-PHA in tumor tissue have proposed functions in the skeletal system and a gene encoding one of these proteins, periostin (POSTN), has been identified as a gene expressed in the myoepithelial cells of breast tumors (Grigoriadis et al., 2006. *Breast Cancer Res* 8:R56). This gene has also been reported to be expressed in normal bone chondrocytes and pre-osteoblasts; both are mesenchymal cell types (Blumer et al., 2006. *J Anat* 208:695-707). We find in our analysis of breast tissue after L-PHA fractionation that we detect POSTN almost exclusively in the tumor. This is the first identification of a difference in the glycosylation of POSTN for breast cancer tissue relative to normal breast tissue. Prior to L-PHA fractionation we detect POSTN in both normal and tumor tissue, likely due to stromal expression. We do not separate epithelial and stromal cells before L-PHA fractionation, therefore, our data strongly support the notion that POSTN with complex $\beta(1,6)$ branched N-linked glycan structures is expressed mainly in the breast tumor epithelial cells since there is very little L-PHA reactive POSTN isolated from normal breast tissue. The activation of POSTN expression in osteoblasts has been linked to twist, a bHLH transcription factor that controls the expression of embryonic mesenchymal genes during development (Oshima et al., 2002. *J Cell Biochem* 86:792-804). Twist expression has

been reported to be increased in lobular infiltrating breast cancer (Yang et al., 2004. *Cell* 117:927-39; Kang and Massague, 2004. *Cell* 118:277-9). Therefore, these observations may explain how POSTN is expressed in breast epithelial cells. Also, TGF β has been shown to increase the expression of POSTN in cardiac development providing another possible explanation for how POSTN may be expressed in breast cancer epithelial cells (Norris et al., 2004. *Anat Rec A Discov Mol Cell Evol Biol* 281:1227-33).

GnT-V, the gene that adds the β (1,6) GlcNAc branch leading to the formation of β (1,6) branched N-linked glycans has recently been implicated in the maintenance of bone density as GnT-V (-/-) mice show a loss of bone mineral density (Cheung et al., 2007. *Glycobiology* 17:828-37). It is possible that POSTN may be a preferred substrate for GnT-V in osteoblasts and chondrocytes during development. Therefore, the abnormal expression of both GnT-V and POSTN in breast carcinoma epithelial cells would lead to significant changes in breast epithelial cell adhesion and migration, promoting tumor progression.

Another protein, osteoglycin (OGN), which was enriched by L-PHA in breast tumor tissue relative to normal tissue was originally identified in bone as an osteoinductive factor (Bentz et al., 1989. *J Biol Chem* 264:20805-10). OGN along with DCN and BGN are members of a group of small leucine-rich repeat proteoglycans (SLRPs) that are important during skeletal development. BGN and DCN are also involved in the development and maintenance of bone. More pronounced loss of bone mass is present in the double knockout of BGN and DCN than for each individual gene knockout suggesting that both of these SLRPs play a role in the maintenance of bone (Corsi et al., 2002. *J Bone Miner Res* 17:1180-9). Interestingly, DCN is also found to be expressed in the myoepithelial cells of the breast (Grigoriadis et al., 2006. *Breast Cancer Res* 8:R56). Exactly how the presence of β (1,6) branched N-linked glycosylation on these proteins within the breast tissue may influence malignancy is unknown. One can postulate that increased β (1,6) branched N-linked glycans may influence the formation of collagen fibrils making it easier for tumor cells to migrate and invade through the basement membrane. Many proteins that we have identified as highly L-PHA reactive in tumor tissue relative to normal tissue have reported functions in skeletal development. This suggests that L-PHA reactive N-linked structures may be promoting mesenchymal functions within breast epithelial cells. Future experiments will test if GnT-V expression levels can affect the epithelial to mesenchymal transition (EMT).

Example III

Focused Glycomic Analysis of the *N*-linked Glycan Biosynthetic Pathway
in Ovarian Cancer

Epithelial ovarian cancer is the deadliest female reproductive tract malignancy in Western countries. Less than 25% of cases are diagnosed when the cancer is confined, pointing to the critical need for early diagnostics for ovarian cancer. Identifying the changes that occur in the glycome of ovarian cancer cells may provide an avenue to develop a new generation of potential biomarkers for early detection of this disease. Epithelial ovarian cancers are comprised of five major subtypes (serous, endometrioid, mucinous, clear cell, and transitional adenocarcinomas), with serous and endometrioid being the two most common types. Epithelial ovarian cancer arises in humans from the ovarian surface epithelium (OSE), epithelial inclusion cysts, or the tubal fimbria (Lee et al., 2007. *J. Pathol.* 211:26–35; Auersperg et al., 2001. *Endocr. Rev.* 22:255–288). Several oncogenes have been implicated in ovarian cancer development including: *c-myc*, *k-ras*, *erbB2*, *egfr*, *p53*, *b-catenin*, *brca1/2*, *pten*, and others (Orsulic et al., 2002. *Cancer Cell* 1:53–62; Aunoble et al., 2000. *Int. J. Oncol.* 16:567–576; Dinulescu et al., 2005. *Nat. Med.* 11:63–70). Utilizing this knowledge, a mouse model of human epithelial endometrioid ovarian carcinoma was developed by adenoviral infection of the Cre recombinase in the OSE of conditional mice engineered to activate oncogenic *k-ras* and inactivate *pten* (Dinulescu et al., 2005. *Nat. Med.* 11:63–70). The combination of these mutations leads to the induction of malignant epithelial endometrioid ovarian carcinoma lesions that recapitulate the morphology and histology of the human disease (Dinulescu et al., 2005. *Nat. Med.* 11:63–70). Several clinical studies have recently validated these genetic results by identifying the first cases of human ovarian carcinomas with synchronous *k-ras* and *pten* mutations (Irving et al., 2005. *Hum. Pathol.* 36:605–619; Kolasa et al., 2006. *Gynecol. Oncol.* 103:692–697).

We performed a glycotranscriptomic analysis of endometrioid ovarian carcinoma using human tissue, as well as a newly developed mouse model that mimics this disease. Our results show that the *N*-linked glycans expressed in both nondiseased mouse and human ovarian tissues are similar; moreover, malignant changes in the expression of *N*-linked glycans in both mouse and human endometrioid ovarian carcinoma are

qualitatively similar. We have used a quantitative real-time PCR approach (qRT-PCR) to quantitatively measure changes in the expression levels of enzymes in the *N*-linked biosynthetic pathway for mouse and human epithelial ovarian endometrioid carcinomas. Comparative lectin blot analysis was used to confirm changes in glycan structures

5 predicted by qRT-PCR results. Lectin reactivity was used as a means for rapid validation of glycan structural changes in the carcinomas that were predicted by the glycotranscriptome analysis. Among several changes in glycan expression noted, the increase of bisected *N*-linked glycans and the transcripts of the enzyme responsible for its biosynthesis, GnT-III, was the most significant. This study provides evidence that
10 glycotranscriptome analysis can be an important tool in identifying potential cancer biomarkers (Abbott et al., August 2008 *Proteomics* 8(16):3210-3220; Abbott et al., August 2008 *Proteomics* 8(16):3210-3220 online Supporting Information available at http://www.wiley-vch.de/contents/jc_2120/2008/pro200800157_s.pdf.

15 Materials and methods

Tumor samples. Endometrioid ovarian tumors ($N = 3$) and normal ovaries ($N = 6$) were obtained from the previously described mouse model (Dinulescu et al., 2005. *Nat. Med.* 11:63–70). Tumors were graded and assessed based on established histopathology analysis (Dinulescu et al., 2005. *Nat. Med.* 11:63–70). Human endometrioid ovarian
20 cancers ($N = 3$) (.90% tumor) were obtained from women as frozen tissue from the Ovarian Cancer Institute (Atlanta, GA). Institutional Review Board approval was obtained for this research. Human normal adjacent ovary RNA samples ($N = 2$) were purchased from BioChain Institute (Hayward, CA). Human normal ovary tissue lysates was purchased from Protein Biotechnologies (Ramona, CA).

25 *qRT-PCR.* Samples (50 mg tissue for tumor, entire normal ovaries) were extracted using 0.8 mL TriZol (Invitrogen, Carlsbad, CA) with polytron homogenization at setting 3 for 1 min. Total RNA was isolated according to the manufacturer's instructions. After DNase treatment, RNA (2 mg) was reverse-transcribed using Superscript III (Invitrogen) with random hexamers and Oligo (dT). Primer pairs for assay genes and control genes
30 were designed within a single exon using conditions described in Nairn *et al.* (2008. *J. Biol. Chem.* 283:17298–17313) and listed in Table 6. Primers were validated with respect to primer efficiency and single product detection. The control gene, Ribosomal Protein L4 (RPL4, NM_024212) was included on each plate to control for run variation and to normalize individual gene expression. Samples were run with negative control templates

prepared without reverse transcription to ensure amplification is specific to cDNA. Triplicate Ct values for each gene were averaged and the SD from the mean was calculated. Data were converted to linear values and normalized as described previously (Nairn et al., 2007. *J. Proteome Res.* 6:4374–4387).

Table 6. Primer pairs for qRT-PCR assay genes and control genes.

Gene	Abbrev.	Species	Accession #	SEQ ID		SEQ ID	
				Forward Primer (5' to 3')	Reverse Primer (5' to 3')	NO:	NO:
Mannoside acetylglucosaminyltransferase 1	Mgat1	mouse	NM_010794	CCCTTCACCCAGTTGGACCTG	GCACCATAGACCTGGCGGAG	308	328
		human	NM_002406	GTGATTCCCATCCTGGTCAT	TAATGCAGCAGCTTGTCCAG	309	329
Mannoside acetylglucosaminyltransferase 2	Mgat2	mouse	NM_146035	TGCTGGAGACTGTGGTATGC	ACTCAATTTGGGCACTCTGG	310	330
		human	NM_002408	AGCAAGAGTGCCCTGAATGT	TGCCATAGAAACTGCGGACTG	311	331
Mannoside acetylglucosaminyltransferase 3	Mgat3	mouse	NM_010795	GCGTGATGGTGTGCTGTTCC	ACAGGGACTTCGCGATGTGG	312	332
		human	NM_002409	CGAGGGCATCTACTTCAAGC	CGCTTGCTCGTAGTCACC	313	333
Mannoside acetylglucosaminyltransferase 4, isoenzyme A	Mgat4a	mouse	NM_173870	GCGACACACAGAAAGGCAAAACC	CCGACAGAGACGAGTGTAGGC	314	334
		human	NM_012214	AAGGTCTACCAAGGGCATACG	TATCGGTGTGATAGCCAGAA	315	335
Mannoside acetylglucosaminyltransferase 4, isoenzyme B	Mgat4b	mouse	NM_145926	AGGTGACGTGGTGGACATTT	GCTTCAGGCTCTCTTGCTCA	316	336
		human	NM_014275	TGCACTCGTACCTGACTGACA	GACCGAGTCTCTCTCTCCT	317	337
Mannoside N-acetylglucosaminyltransferase V	Mgat5	mouse	NM_145128	CCCTGGAAGTTGTCCTCTCA	TCCTCTGCCAGTGCCTTAAT	318	338
		human	NM_002410	AGCCTGAAAGCAGCTCCAT	GCCAGTGCCTTGATGTACCT	319	339
Mannosidase 2, alpha 1 (Man2a1)	Man1l	mouse	NM_008549	CTCTGGTGTCGCTGCTGGTG	AATAGGCAGGACTGGCGAACC	320	340
		human	NM_002372	GAGTGAGTCTGTGGAGGATGG	GAAGATGTAGCCAGCACCT	321	341
Mannosidase 2, alpha 2 (Man2a2)	Man1lx	mouse	NM_172903	AAGCAGGTGACGGTGTGTGG	TGCACCTGGTCCAGCATAAGG	322	342
		human	NM_006122	GCCCAGCTTCTTCCATCT	TGCAGCTGGCTTCTGAC	323	343
Fucosyltransferase 8	Fut8	mouse	NM_016893	AACAGAAGCAGCCTTCCACCC	CATTCTGGTGGCGAGAAGCTG	324	344
		human	NM_178154	GGTCGAGCTTCCATTGTAG	GCGAGGTCTTCTGTGACAGC	325	345
Ribosomal Protein L4*	RPL4	mouse	NM_024212	GACAGCCCTATGCCGTGAGTG	GCCACAGCTCTGCCAGTACC	326	346
		human	NM_000968	AGAAGGCTGCTGTGGTGT	TGGTTCTTGGTAGCTGCTG	327	347

*Normalization control gene

Lectin analysis. Tissue (50 mg) isolated from the same tumors used for qRT-PCR analysis were lysed in RIPA buffer (16PBS, 1% NP-40, 0.5% DOC, 0.1% SDS) containing a mini complete protease inhibitor tablet (Roche, Indianapolis, IN) using a polytron at setting 3 for 1 min. The lysate was cleared by centrifugation at 10 000 x g for 10 min. Protein concentrations were determined by BCA assay (Pierce, Rockford, IL). Biotinylated lectin (Con A, Vector Labs, Burlingame, CA) (2 mg) was added to 50 mg cleared lysate at 47°C for 2 h. Paramagnetic streptavidin beads (50 mL) were added to separate Con A bound and unbound proteins. Unbound fractions, 10 mg, were separated on 4–12 % NuPage Bis Tris gels and transferred to PVDF membrane at 25 V for 1.5 h. Membranes were blocked overnight in 3% BSA/TBST buffer before lectin blot detection using a 1:5000 dilution of the following biotinylated lectins: (*Phaseolus vulgaris* leucoagglutinin (L-PHA), *P. vulgaris* erythroagglutinin (E-PHA), *Aleuria aurantia* (AAL), and *Datura stramonium* (DSL), Vector Labs). Bound lectin was detected using a 1:5000 dilution of streptavidin–HRP (Vector Labs) before washing and detection using Western Lightening Plus (Perkin Elmer).

Results

Expression data from the N-linked biosynthetic pathway for normal ovary. Enzymes from the N-linked glycosylation pathway (Fig. 8A) were chosen for analysis based on low levels of redundancy and the previous correlations of transcript levels with glycan structural analysis in mouse tissues (Nairn et al., 2008. *J. Biol. Chem.* 283:17298–17313). Before analyzing tumor tissues, we examined the relative transcript abundance of these enzymes using pooled RNA isolated from normal mouse and human ovarian tissue. The development of epithelial endometrioid ovarian tumors in the model developed by Dinulescu *et al.* (2005. *Nat. Med.* 11:63–70) utilizes adenoviral infection of the Cre recombinase into the outer epithelial layer of the ovary to initiate tumor formation. This technique allows the noninfected ovary within the same animal to serve as a normal control. From the group of genes analyzed, there are both similarities and differences in transcript expression profiles between mouse and human normal ovary. Transcripts present in higher abundance for both species are: FUT8, MGAT1 (human) or MGAT2 (mouse), and MGAT5 (Figs. 8B and C). These enzymes are involved in the branching (MGAT1, MGAT2, MGAT5) or core fucosylation (FUT8) of N-linked glycans (Fig. 8A). Due to the higher levels of these

transcripts, core fucosylated hybrid type or core fucosylated complex branched *N*-linked glycans are likely to be abundant in both mouse and human ovary. The MGAT3 transcripts are present at low levels in both mouse and human ovaries (Figs. 8B and C). The lower levels of MGAT3 transcripts suggest that bisecting *N*-linked glycans may be present at lower levels in normal mouse and human ovary tissue. Interestingly, the levels of MGAT4b transcripts are very high in mouse compared with human. However, for both mouse and human the MGAT4a levels are lower. These genes are both capable of adding *N*-acetylglucosamine in $\beta(1,4)$ linkage (Fig. 8A). Considering the levels of both MGAT4a and MGAT4b, along with abundant MGAT5 levels, tri- and tetra-antennary complex *N*-linked glycans should be present in normal ovary. Enzymes showing differences in transcript abundance in mouse and human are the mannosidases, MAN2A1 (Man II) is more abundant in human ovary, while MAN2A1 (Man II) and MAN2A2 (Man IIx) are equally abundant in mouse ovary. Overall, for this subset of genes participating in the *N*-linked pathway, transcript levels in both mouse and human normal ovary show a high degree of species conservation.

Comparative analysis of normal and epithelial endometrioid ovarian carcinoma. To investigate possible differences in the expression of GT and GH in malignant epithelial ovarian tissue, RNA from mouse endometrioid and human endometrioid carcinoma was analyzed by qRT-PCR. Total RNA from age-matched human normal ovary was purchased from a commercial source. Human normal ovarian tissue samples were averaged for comparison with the qRT-PCR results from individual human tumor tissues. The mouse normal samples qRT-PCR results were also averaged to enable comparison with qRT-PCR results from individual mouse tumor tissues. The levels of MGAT1 and MGAT2 transcripts were increased above normal for both mouse and human tumors suggesting increased complex *N*-linked glycans (Figs. 9A and B). The transcripts encoding FUT8, the enzyme responsible for $\alpha(1,6)$ fucosylation of the core *N*-linked glycan, increased on average 1.5–2-fold relative to normal for mouse tumor tissues (Fig. 9A). However, in human endometrioid ovarian tumors, an increase in FUT8 transcripts occurred in only one of three cases analyzed (Fig. 9B). These results indicate that the factors regulating transcript levels of FUT8 may be more complex in human ovarian cancer. The transcript levels for MGAT3 in both mouse and human tumor samples were increased significantly (average of 18-fold for human and 16-fold for mouse) relative to normal ovarian tissue. The lower transcript abundance of MGAT3

observed for mouse and human normal ovary (Figs. 8B and C) contrast with the large increase in MGAT3 transcripts observed for all cases of epithelial endometrioid carcinoma analyzed in both mouse and human (Figs. 9A and B). The commonality of this change for mouse and human tumor samples along with the magnitude of the change predict the possibility of isolating glycoproteins with complex bisecting *N*-linked glycans as markers for endometrioid ovarian cancer. Transcripts encoding enzymes that perform outer branching of complex *N*-linked glycans such as MGAT4a, MGAT4b, and MGAT5 are increased in mouse and human ovarian cancer relative to normal. MGAT4a transcripts are increased at a higher level relative to normal for mouse compared with human (2–6-fold and 1.5–4-fold, respectively). However, levels of MGAT4b were also elevated in human tumor tissue, while mouse MGAT4b transcript levels in mouse tumor were not significantly increased. The different regulation of MGAT4b transcripts in mouse and human ovarian tumor tissue may highlight a possible species-specific difference in the transcriptional regulation of this gene. However, the cumulative effect of increases in MGAT4a, MGAT4b, and MGAT5 predict more branched complex *N*-linked glycan structures present in endometrioid ovarian cancer relative to normal.

Comparative lectin analysis. To investigate whether the tumor-specific changes in GT transcript levels correlate with glycan structures found on glycoproteins, we used lectin separation and blotting techniques. The carbohydrate binding preferences for lectins are diverse, allowing the selection of lectins to detect a wide range of oligosaccharide structures. Lectins (E-PHA, LPHA, DSL, and AAL) recognizing the oligosaccharide products of FUT8, MGAT3, MGAT4a, MGAT4b, MGAT5, and MGAT5b are shown in Fig. 10. These are only examples of oligosaccharide structures that these lectins can bind to and are not intended to be a complete list of all structures capable of binding the lectins. Sugar residues previously described as a determinant for binding of each lectin are circled (Nagata et al., 1991. *Biochim. Biophys. Acta* 1076:187–190; Wimmerova et al., 2003. *J. Biol. Chem.* 278:27059–27067; Cummings and Kornfeld, 1982. *J. Biol. Chem.* 257:11235–11240; Cummings and Kornfeld, 1982. *J. Biol. Chem.* 257:11230–11234; Yamashita et al., 1987. *J. Biol. Chem.* 262:1602–1607). Con A is a lectin that recognizes branched mannose residues with high affinity (Fig. 10), and this lectin was used to separate the high-mannose, hybrid, and complex biantennary oligosaccharides from the complex tri- and tetra-antennary *N*-linked glycans

prior to lectin blot detection (Cummings and Kornfeld, 1982. *J. Biol. Chem.* 257:11235–11240). Total cell lysates from the three mouse tumors were pooled for lectin analysis due to the high degree of correspondent changes observed in the qRT-PCR experiments.

Increased core fucosylation in ovarian tumors. All three mouse endometrioid ovarian
5 tumors analyzed showed elevated levels of FUT8 transcripts (Fig. 9). The lectin AAL has a high affinity for the core $\alpha(1,6)$ fucose linked product that would result from FUT8 activity (Nagata et al., 1991. *Biochim. Biophys. Acta* 1076:187–190). Core fucosylation can be found on hybrid type *N*-linked glycans as well as complex bi-, tri-, and tetra-antennary oligosaccharides. However, no differences in AAL binding to Con A bound fractions were
10 observed between normal and tumors, suggesting that the levels of core fucosylation do not change significantly on hybrid type and complex biantennary *N*-linked glycans for ovarian tumors (data not shown). AAL binding to the unbound Con A fraction increases significantly to glycoproteins isolated from ovarian tumors compared to normal ovarian tissue (Fig. 11A lanes 1 and 2). These results indicate that the core fucosylation of complex tri- and tetra-
15 branched oligosaccharides increased for tumors relative to normal. To analyze human ovarian cancer tissues, individual samples were analyzed due to the differences in qRT-PCR results for these cases for FUT8. AAL binding to glycoproteins isolated from human endometrioid ovarian tumors was similar to the levels predicted by qRT-PCR results for case 471. Case 471 had a .4-fold increase in FUT8 mRNA compared with normal and shows
20 seven-fold increase in the level of AAL binding relative to normal (Figs. 9B and 11B, lane 4). Case 711 had similar levels of FUT8 transcripts and shows levels of AAL reactivity similar to the normal control sample (Figs. 9B and 11B, lane 2). Case 741 had lower levels of FUT8 expression compared with normal (Fig. 9B), yet AAL binding was significantly increased (Fig. 11B, lane 3). To better evaluate the fucosylation of this case we used qRT-
25 PCR to analyze all known fucosyltransferases and fucosidases. Our results indicate a two-fold reduction in the levels of α -L-fucosidase (FUCA2) expression (data not shown). We did not observe this decrease in FUTA2 for cases 711 and 471, suggesting that this may be contributing to the increased levels of core α -1,6-fucosylation for case 741. Our data suggest that in the majority of cases for both mouse and human, core fucosylation levels are elevated.
30 Therefore, glycoproteomic studies targeting the core fucose could potentially lead to the isolation of tumor-specific markers for ovarian cancer.

Ovarian tumors have increased levels of bisecting complex N-linked glycans. N-glycans containing a bisecting N-acetylglucosamine are produced by the activity of MGAT3 (Fig. 8A). This enzyme showed the largest increases in expression in ovarian tumor tissues compared with normal ovary tissue (Figs. 9A and B). Lectin analysis of mouse tissues using the lectin E-PHA, whose binding is dependent on the presence of bisecting GlcNAc (Fig. 10), shows .2X the levels of E-PHA binding for tumor relative to normal (Fig. 11A, lanes 3 and 4). Human ovarian tumor glycoproteins analyzed for E-PHA binding showed positive correlation with qRT-PCR results for each case. For example, case 471 and 741 had the highest and second highest increases in MGAT3 transcript levels (Fig. 9B) and these cases also have the highest levels of EPHA binding (Fig. 11B, lanes 7 and 8). Case 711 had a lower increase in MGAT3 transcripts measured by qRT-PCR and shows a lower level of E-PHA binding than cases 471 and 741. In every endometrioid ovarian tumor analyzed, mouse and human, there were elevated bisecting complex N-linked glycans compared with normal controls. Although the datasets are small, these results strongly suggest that the presence of bisecting complex N-glycans is a marker for ovarian tumors.

Tri- or tetra-antennary complex N-linked glycans are increased in ovarian tumors.

The lectins known as DSL and L-PHA recognize the outer branches of complex N-linked glycans (MGAT4a, MGAT4b, MGAT5, and MGAT5b) (Fig. 10). L-PHA is specific for the $\beta(1,6)$ branched, galactosylated product of MGAT5; while DSL can recognize the $\beta(1,4)$ branch added by MGAT4a and MGAT4b as well as the MGAT5 branch (Fig. 10). MGAT4a and MGAT4b transcript levels were elevated for human endometrioid ovarian tumors and MGAT4a transcripts were amplified for mouse tumors (Figs. 9A and B). These genes can perform the same glycosyltransferase reactions and their elevations predict the existence of more $\beta(1,4)$ branched glycans and increased DSL binding. Indeed, analysis of unbound Con A fractions indicate elevated levels of DSL binding to mouse tumor glycoproteins compared with glycoproteins from normal mouse ovary (Fig. 11A, lanes 5 and 6). The levels of DSL binding to tumor glycoproteins isolated from human ovarian tumors were lower compared with AAL or E-PHA binding (Fig. 11B, lanes 9–12) and this correlates with a lower change in transcript levels for MGAT4 (Fig. 9B). Interestingly, L-PHA binding levels for mouse ovarian tumor tissues were not increased above normal, despite increases in MGAT5 transcripts (Figs. 9A and 11A, lanes 7 and 8). Human tumor samples had higher levels of

MGAT5 transcripts compared with normal and do show a slight increase in LPHA binding compared with normal (Figs. 9B and 11B, lanes 13–16). Mouse ovarian tumors had a two-fold increase in MGAT5 and human ovarian tumors had a three- to five-fold increase in MGAT5 expression (Figs. 9A and B and Table 7). Based on these findings elevated L-PHA binding would be expected. However, we find no increase in L-PHA binding for mouse ovarian tumors (Fig. 11A) and a lower than expected increase in L-PHA reactivity to glycoproteins from human ovarian tumors. These results may indicate that the MGAT3 elevated activity may inhibit MGAT5 activity and this will be discussed further in the discussion. Overall, the DSL lectin is useful for capturing glycoproteins with the MGAT4 or MGAT5 branch, and based on our analysis, would be an effective lectin for isolating glycoprotein markers with elevated β (1,4) or β (1,6) complex branched glycans for endometrioid ovarian tumors.

Table 7. Genes analyzed by qRT-PCR.

Symbol	Average fold change mouse	Average fold change human	Chromosome location	Mouse	Human
Fut8	(+) 1.5	(+) 2.3	14q24.3	NM_016893	NM_178154
Mgat1	(+) 2.0	(+) 5.2	5q31 or 5q35	NM_010794	NM_002406
Mgat2	(+) 1.9	(+) 3.8	14q21	NM_146035	NM_002408
Mgat3	(+) 16.0	(+) 18.2	22q13.1	NM_010795	NM_002409
Mgat4a	(+) 4.4	(+) 2.5	2q12	NM_173870	NM_012214
Mgat4b	no change	(+) 1.5	5q35	NM_145926	NM_014275/NM_054013
Mgat5	(+) 2.2	(+) 4.0	2q21	NM_145128	NM_002410
MAN2A1	(+) 2.5	(+) 1.4	5q21	NM_008549	NM_002372
MAN2A2	(+) 1.9	(+) 1.2	15q25	NM_172903	NM_006122

Discussion

In this report we have used a glycotranscriptome approach to characterize the *N*-linked glycan profiles of normal ovary and endometrioid ovarian carcinoma. Our results provide several significant findings: (i) mouse and human normal ovarian tissues have a similar expression profile for certain enzymes participating in the formation of *N*-linked glycans suggesting some degree of species conservation, (ii) enzymes changing in expression

for tumors isolated from the mouse model of human endometrioid ovarian cancer correspond qualitatively with changes observed for human tumors of the same malignancy, (iii) E-PHA, AAL, and DSL reactivity levels were elevated in endometrioid ovarian tumors relative to normal indicating that these lectins could be useful together for biomarker discovery or to improve the specificity of existing ovarian tumor markers.

Glycosylation changes observed for reproductive malignancies. Analysis of glycoprotein glycosylation patterns is emerging as a powerful tool to discriminate normal glycoproteins from glycoproteins marking diseases such as cancer. Over 20 years ago, researchers began to observe glycosylation changes occurring on cell surface glycoproteins following oncogenic transformation (Pierce and Arango, 1986. *J. Biol. Chem.* 261:10772–10777; Meezan et al., 1969. *Biochemistry* 8:2518–2524; Kobata and Amano, 2005. *Immunol. Cell Biol.* 83:429–439; Hakomori, 1999. *Biochim. Biophys. Acta* 1473:247–266). For breast cancer, increased $\beta(1,6)$ branched glycans emerged qualitatively from the transition to malignancy due to elevated expression of the MGAT5 gene (Buckhaults et al., 1997. *J. Biol. Chem.* 272:19575–19581; Fernandes et al., 1991. *Cancer Res.* 51:718–723). Few studies have evaluated glycosylation changes for ovarian cancer (An et al., 2006. *J. Proteome Res.* 5:1626–1635; Saldova et al., 2007. *Glycobiology* 17:1344–1356.; Kui Wong et al., 2003. *J. Biol. Chem.* 278:28619–28634; Wang et al., 2005. *Gynecol. Oncol.* 99:631–6390). However, from the small group of studies conducted, Rudd and coworkers (Saldova et al., 2007. *Glycobiology* 17:1344–1356) found increased levels of bisecting core fucosylated complex *N*-linked glycans in the serum of ovarian cancer patients compared with normal serum. This result agrees with our data showing increased core fucosylation and bisecting glycans in endometrioid ovarian tumor tissues. Therefore, there is a high probability of isolating and identifying glycoproteins shed into serum from ovarian carcinomas with differences in glycosylation. Lebrilla (An et al., 2006. *J. Proteome Res.* 5:1626–1635) have monitored changes in glycan structures in the serum of ovarian cancer patients using MALDI-FTMS and find several tumor-specific changes for neutral oligosaccharides in MALDI spectra. This study employed a β -elimination procedure to remove oligosaccharides; therefore the neutral glycans changing in the serum could correspond to *N*- or *O*-linked glycans. Our data suggest that neutral glycans from the *N*-linked biosynthetic pathway such as core fucosylation and bisecting *N*-acetylglucosamine are increased significantly in ovarian tumor tissue relative to

normal. Therefore, it should be possible to identify the glycoproteins that have these neutral *N*-linked glycan alterations from patient serum. In conclusion, our data in addition to these previous studies indicate that like breast cancer, ovarian tumor formation results in distinct altered glycan structures.

5 *Comparison of changes in glycosylation between a mouse-model of endometrioid ovarian cancer and human endometrioid ovarian cancer.* Ovarian tumors derived from the mouse model of human endometrioid ovarian cancer developed by Dinulescu *et al.* (2005. *Nat. Med.* 11:63–70) are well differentiated and recapitulate human ovarian cancer histologically. The benefits of mouse models are numerous, including: stable genetics,
10 controlled environmental factors such as diet, and most importantly, the ability to sample tissue and serum at different stages of tumor development. Our data demonstrate that the glycomic changes occurring in the *N*-linked pathway for mouse-derived epithelial endometrioid ovarian tumors show corresponding changes in human ovarian endometrioid tumors.

15 *Increased knowledge about the N-linked glycan pathway.* The *N*-linked glycosylation pathway consists of a series of sequential reactions (Fig. 8A). The qRT-PCR approach enabled the quantization of the transcript levels for several enzymes with a wide dynamic range. The total pathway approach offers a chance to learn more about how synchronous oncogenic signaling changes can influence glycosylation. A summary of the average fold
20 changes in expression measured for each enzyme analyzed is provided in Table 7. Some interesting findings that have resulted from this study include: (i) increased variability in FUT8 levels for human ovarian cancer versus mouse tumors, suggesting more complex influences controlling the levels of core fucosylation, (ii) lower levels of L-PHA reactivity for glycoproteins with elevated MGAT5 mRNA levels suggests possible inhibition of
25 MGAT5 activity by MGAT3.

Genetic factors influencing glycosylation patterns have not been extensively studied. Although there has been a study recently published examining the effect of single nucleotide polymorphisms in genes involved in the mucin-type glycosylation of MUC1 (Sellers *et al.*, 2008. *Cancer Epidemiol. Biomarkers Prev.* 17:397–404). This study found that genetic
30 polymorphisms within glycosylation enzymes analyzed for MUC1 may be playing a role in the underglycosylation of this protein in ovarian cancer patients suggesting that genetic

factors can affect glycosylation. Several of the GT and GH enzymes included in this study are located in similar regions of the chromosomes (Table 7). MGAT1 and MGAT4b are located on the same chromosome in close proximity and show very different expression profiles for ovarian cancer. This seems to suggest that elevations in MGAT1 expression are
5 probably not related to gain of chromosome copy, or MGAT4b would be increased in a similar manner. The changes in expression observed for GT enzymes in ovarian cancer could be due to differences in factors regulating GT promoters.

The small sample size of human endometrioid carcinoma cases analyzed in our study suggests that FUT8 expression and activity are more variable. Due to the lack of variability
10 in the mouse model, we postulate that in humans there may be unknown factors in ovarian tumors either genetic or epigenetic that are capable of influencing core fucosylation. The FUT8 variability in human ovarian tumors contrasts with MGAT1, MGAT2, and MGAT3 which seems to be unaffected by genetic differences. More studies examining the glycomic changes in human cancer samples performed in conjunction with murine models are needed
15 to better understand the possible role of genetic regulation on glycosylation.

The product of MGAT3 activity, the bisecting *N*-acetylglucosamine structure, has been reported to inhibit the activity of MGAT5 (Yoshimura et al., 1995. *Proc. Natl. Acad. Sci. USA* 92:8754–8758; Taniguchi et al., 1996. *Glycobiology* 6:691–694). We find that MGAT3 levels are substantially increased far above MGAT5 levels in ovarian tumors.
20 Therefore, one possibility for why we observe less change in L-PHA reactivity for ovarian tumors (Figs. 11A and B), despite increased levels of MGAT5 expression, may be that the addition of bisecting *N*-acetylglucosamine by MGAT3 prevented the addition of the $\beta(1,6)$ branched oligosaccharide structure. Competition for the same nucleotide sugar donor (UDP-GlcNAc) does not seem to be a factor since MGAT4a and MGAT4b addition (evidenced by
25 DSL binding) is unaffected by the large increase in MGAT3 activity. The addition of bisecting bisecting *N*-acetylglucosamine to the trimannosyl core *N*-linked glycan has been reported to occur in opposition to $\beta(1,6)$ branching performed by MGAT5 during the cell cycle (Guo et al., 2000. *Biochim. Biophys. Acta* 1495:297–307). In this study the authors found that the mRNA levels and protein levels for MGAT5 were not changing, yet there was less MGAT5
30 enzyme activity at stages of the cell cycle when MGAT3 activity levels were high. These data along with our data support the notion that increased MGAT3 activity inhibits the

addition of $\beta(1,6)$ branched glycans by MGAT5 and the mechanism of this inhibition is currently unknown.

Bisecting oligosaccharides and cancer. Elevated levels of MGAT3 expression and an increase in bisecting glycans have been reported for pancreatic cancer and hepatoma (Ishibashi et al., 1989. *Clin. Chim. Acta* 185:325–332; Nan et al., 1998. *Glycoconj. J.* 15:1033–1037). Studies using diethylnitrosamine to induce liver tumor formation in mice null for MGAT3 showed reduced tumor formation (Wang et al., 2005. *Gynecol. Oncol.* 99:631–639; Bhaumik et al., 1998. *Cancer Res.* 58:2881–2887). However, if MGAT3 was overexpressed there is no induction or augmentation of tumor growth using diethylnitrosamine (Stanley, 2002. *Biochim. Biophys. Acta* 1573:363–368). This result suggests an indirect effect of MGAT3 overexpression on liver tumorigenesis. In ovarian tumors, bisecting glycan structures predominate as evidenced by increased MGAT3 mRNA (Figs. 9A and B) and elevated EPHA binding (Figs. 11A and B). Bisecting structures have also been documented on CA125 isolated from the ovarian cancer cell line OVCAR3 (Kui Wong et al., 2003. *J. Biol. Chem.* 278:28619–28634). Mice null for MGAT3 are viable and reproduce normally. This, along with the fact that MGAT3 is expressed at a low level in normal ovary, suggest that therapeutic strategies targeting MGAT3 could be useful for retarding ovarian tumor progression with minimal interruption of normal ovary functions.

In conclusion, the glycomic analysis presented in this manuscript provides a framework for future glycoproteomic studies. These studies will enable the identification of the proteins that express bisecting *N*-linked glycans and allow for the structural characterization of the bisecting oligosaccharides.

Example IV
Ovarian Cancer Biomarkers

5 Glycan changes present in ovarian cancer may differ from glycan changes present in breast cancer. Potential biomarkers of ovarian cancer, as shown in Table 8, were discovered in ovarian tissue using methods analogous to those described in Examples II and III, using lectins that recognize bisecting *N*-linked glycans (E-PHA), core-fucosylated (AAL), and $\beta(1,4)$ branched glycans (DSL). As shown below, not all of the identified proteins were predicted to be glycosylated proteins (glycoproteins). These may 10 associate with other, lectin-reactive glycoproteins or have acquired a glycan to become a glycoprotein as part of oncogenic transformation.

15 Table 8. Lectin-reactive proteins with elevated peptides and spectra in ovarian tumor relative to normal.

Name	Tumor Total	Normal Total	Predicted Glycoprotein
Glucose-regulated protein 78	16	0	No
heterogeneous nuclear ribonucleoprotein F	7	0	No
lysosomal-associated membrane protein 1	8	0	Yes
periostin	20	5	Yes
adenylate cyclase-associated protein 1 (yeast)	7	0	No
protein disulfide isomerase family A, member 4	4	0	No
chaperonin containing TCP1, subunit 5 (epsilon)	8	0	No
biglycan	22	8	Yes
procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase)	14	0	No
Ceruloplasmin	8	0	Yes
transforming growth factor, beta-induced	16	6	No
myosin, heavy chain 9, non-muscle	34	7	No
complement factor B	7	0	Yes
myosin, heavy chain 11, smooth muscle	21	5	No
calreticulin	14	3	No
creatin kinase-brain	7	0	No
lectin, galactoside-binding, soluble, 3 binding protein	8	0	Yes
heparan sulfate proteoglycan 2	7	0	Yes
myosin, heavy chain 11, smooth muscle	21	5	No
protein disulfide isomerase family A, member 3	16	0	No
tenascin XB	35	28	Yes
heat shock 27kDa protein 1	13	4	No
glucose-6-phosphate isomerase	7	0	No
phosphoglycerate kinase 1	16	3	No
elongation factor 2	10	3	No

Name	Tumor Total	Normal Total	Predicted Glycoprotein
ADP-Ribosylation Factor 1	5	0	No
ADP-Ribosylation Factor 3	5	0	No
heterogeneous nuclear ribonucleoprotein K	9	0	No
heterogeneous nuclear ribonucleoprotein K	9	0	No
lactate dehydrogenase A	18	2	Yes
Isoform 2 of Periostin precursor	20	5	Yes
phosphogluconate dehydrogenase	5	2	No
glutathione S-transferase pi 1	6	4	No
peroxiredoxin 6	5	0	No
14-3-3GAMMA	5	0	No
malate dehydrogenase 2, NAD (mitochondrial)	6	0	No
thrombospondin 1	22	0	Yes
coagulation factor XIII, A1 polypeptide	8	0	Yes
chaperonin containing TCP1, subunit 2 (beta)	4	0	No
lactotransferrin	4	0	Yes
talin 1	9	0	No
protein disulfide isomerase family A, member 6	15	0	No
eukaryotic translation initiation factor 5A	5	0	No
fibulin 5	5	2	Yes
immunoglobulin heavy constant gamma 2 (G2m marker)	8	0	Yes
periostin, osteoblast specific factor	20	5	Yes
eukaryotic translation initiation factor 5A	5	0	No
similar to hCG2038920	8	0	Yes
XP_933498 similar to Phosphoglycerate	5	0	No
alpha-2-macroglobulin	31	2	Yes
heterogeneous nuclear ribonucleoprotein U	5	0	No
phosphoglycerate mutase 1 (brain)	5	0	No
transgelin 2	13	2	No
periostin, osteoblast specific factor	20	5	Yes
biglycan	22	8	Yes
heterogeneous nuclear ribonucleoprotein U	5	0	No
heterogeneous nuclear ribonucleoprotein U	5	0	No
protein disulfide isomerase family A, member 6	15	0	No
transgelin 2	13	2	No
protein disulfide isomerase family A, member 3	16	0	No
immunoglobulin lambda locus	13	8	Yes
myosin, heavy chain 11, smooth muscle	21	5	No
myosin, heavy chain 11, smooth muscle	21	5	No
phosphogluconate dehydrogenase	5	2	No
talin 1	9	0	No
mucin 5B	18	0	Yes
lactotransferrin (truncated)	4	0	Yes
lactotransferrin (truncated)	4	0	Yes
lactotransferrin (truncated)	4	0	Yes
glutathione S-transferase pi 1	6	4	No
glutathione S-transferase pi 1	6	4	No
periostin, osteoblast specific factor	20	5	Yes
glutathione S-transferase pi 1	6	4	No
heterogeneous nuclear ribonucleoprotein K	9	0	No

Name	Tumor Total	Normal Total	Predicted Glycoprotein
immunoglobulin heavy constant gamma 3 (G3m marker)	6	0	Yes
serpin H1 precursor (SERPINH1)	20	8	Yes

Example V

Detection of Breast Cancer Biomarkers in Serum

5

Increased $\beta(1,6)$ branched glycans in breast cancer. Serum (5 μ l) from patients with breast cancer (BC) and healthy patients (NL) was incubated with 10 μ g of biotinylated LPHA in 300 μ l of lectin binding buffer (50 mM Tris-Cl pH 7.5, 0.1 % NP-40, 150 mM NaCl, 0.4 mM EDTA, and 1 protease inhibitor tablet) overnight at 4°C. Paramagnetic streptavidin beads were added and the reactions were incubated at room temperature for 1 hour. Bound complexes were captured using a magnet and washed with lectin binding buffer 3X before separating the proteins on a 4-12 % NuPage Bis-Tris gel. Proteins were transferred to PVDF membrane and probed using the indicated antibodies, anti-periostin and anti-osteoglycin. Periostin (POSTN) from breast cancer (BC) cases (BC3, BC4, BC5 and BC6) show increased reactivity with LPHA compared with the POSTN from normal (NL) cases (Fig. 12). These cases also exhibit an additional isoform of periostin reacting with L-PHA that is of a higher molecular weight. Osteoglycin (OGN) from BC cases (BC1, BC3, BC4, BC5, and BC6) also shows increased reactivity with L-PHA compared with OGN from NL cases. Combining these results, serum from BC1, BC3, BC4, BC5, and BC6 have significant increases in L-PHA reactive POSTN and/or OGN compared to serum from healthy women. Fig. 12 thus provides evidence of increased $\beta(1,6)$ branched glycans on periostin (POSTN) and osteoglycin (OGN) (also referenced as mimecan) in serum from breast cancer (BC) cases compared with normal (NL) serum.

25

Example VI

Detection of Ovarian Cancer Biomarkers in Tissue and Serum

Tumor-specific glycosylation in ovarian cancer. Markers were selected for validation based on results obtained using glycoproteomic analysis of endometrioid ovarian tumor (TU)

tissue compared with benign (NL) ovary tissue (Example IV). Lysosomal-associated membrane glycoprotein 1 (LAMP-1) was immunoprecipitated from 500 µg of pre-cleared protein lysate obtained from ovarian tumor tissue or normal ovarian tissue in 300 µl of 1X Tris-buffered saline, 1% triton-X-100 using 2 µg of anti- LAMP-1 monoclonal antibody (E-
5 Bioscience). Bound antibody complexes were captured using 50 µl of protein A/G plus agarose (Santa Cruz Biotechnology) at 4°C for 2 hours. Complexes were washed stringently using 1X TBS/1% triton X-100 buffer before separating the proteins on 4-12 % NuPage Bis-Tris gels prior to transfer to PVDF membrane for 2 hours at 25V. Membranes were blocked overnight at 4°C in 3 % BSA/1X TBST before detection using the indicated biotinylated
10 lectin (DSL, AAL, E-PHA) (1:5,000 dilution) followed by streptavidin-HRP (1:5,000) incubation and Western Lightening Plus (Perkin Elmer) detection. LAMP-1 isolated from tumor tissue shows increased reactivity with the DSL and AAL lectins compared with LAMP-1 isolated from normal ovary (Fig. 13A). The lower band represents a non-specific reaction of the streptavidin-HRP with a protein from the tissue lysates. This band serves as a
15 control for equivalent protein content in the immunoprecipitation reactions; confirming that the increased band density for LAMP-1 from tumor tissue compared with normal is due to increased glycosylation and not a difference in protein input. These results definitively demonstrate significantly elevated levels of branched N-linked glycans that have either α 1,3 or α 1,6 linked fucose on LAMP-1 in ovarian tumor tissue compared to benign ovarian tissue.

20 POSTN was immunoprecipitated from 500 µg of pre-cleared protein lysate as described above using 2 µg of an anti-periostin polyclonal antibody (Abcam). There is no POSTN reacting with the DSL lectin. However, POSTN isolated from tumor tissue is reacting with the AAL and E-PHA lectins, while there is no detectable band present in the POSTN isolated from normal tissue (Fig. 13B). These results suggest that POSTN from ovarian tumor tissue
25 has increased core fucosylation and increased levels of bisecting glycans compared with POSTN from normal ovary. The lower non-specific band again demonstrates an equivalent amount of protein present in the precipitation reactions. Lectin galactoside soluble binding protein 3 (LGALS3BP) was immunoprecipitated as described for LAMP-1 using a goat polyclonal antibody to LGALS3BP (Santa Cruz Biotechnology). LGALS3BP isolated from
30 normal and tumor tissue shows no change in reactivity to DSL and AAL, however, there is an increased reactivity with E-PHA for LGALS3BP from ovarian tumor tissue compared to

normal (Fig. 13C). These results suggest that LGALS3BP has increased bisecting N-linked glycans in ovarian tumor tissue relative to normal ovarian tissue.

Increased bisecting glycans in ovarian cancer. Serum (5 μ l) was incubated with 10 μ g of biotinylated E-PHA in 300 μ l of lectin binding buffer (50 mM Tris-Cl pH 7.5, 0.1 % NP-40, 150 mM NaCl, 0.4 mM EDTA, and 1 protease inhibitor tablet) overnight at 4°C.

Paramagnetic streptavidin beads were added and the reactions were incubated at room temperature for 1 hour. Bound complexes were captured using a magnet and washed with lectin binding buffer 3X before separating the proteins on a 4-12 % NuPage Bis-Tris gel. Proteins were transferred to PVDF membrane and probed using the indicated antibodies, anti-periostin and anti- α -1 acid GP as a control. Ovarian tumors (OT) OT1, OT4, and OT5 have increased levels of periostin interacting with EPHA when compared to all benign cases analyzed (Fig. 14A). These cases also show evidence of multiple isoforms of periostin present after EPHA precipitation. These results indicate that POSTN present in the serum of patients with ovarian cancer have bisecting N-linked glycans while POSTN that may be present in the serum of women with benign gynecologic diseases does not have this glycan structure. These results correlate with the immunoprecipitation results obtained from ovarian cancer tissue compared with normal ovarian tissue (Fig. 13B). α -1 acid glycoprotein (α -1 acid GP) is a ubiquitous protein found in serum that does not show any changes in the abundance of bisecting N-linked glycans (Fig. 14B). These results demonstrate that equivalent levels of protein were present in the E-PHA precipitation reactions. Therefore, the absence of POSTN in the E-PHA precipitation from serum of women with benign gynecologic disease is not due to protein degradation or reduced protein input. These results provide evidence that the presence of bisecting N-linked glycans on periostin is an ovarian tumor-specific marker found in serum.

Increased fucosylated glycans on lysosomal-associated glycoprotein 1 (LAMP-1).

Serum (5 μ l) was incubated with 10 μ g of biotinylated AAL in 300 μ l of lectin binding buffer (50 mM Tris-Cl pH 7.5, 0.1 % NP-40, 150 mM NaCl, 0.4 mM EDTA, and 1 protease inhibitor tablet) overnight at 4°C. Paramagnetic streptavidin beads were added and the reactions were incubated at room temperature for 1 hour. Bound complexes were captured using a magnet and washed with lectin binding buffer 3X before separating the proteins on a 4-12 % NuPage Bis-Tris gel. Proteins were transferred to PVDF membrane and probed

using the indicated antibodies. Ovarian tumors (OT1, OT2, OT3, and OT5) have increased levels of Lamp-1 interacting with AAL compared to benign cases analyzed (Fig. 15). These cases also show evidence of a slower migrating form of Lamp-1, further evidence of a change in glycosylation. These results correlate with the immunoprecipitation results
5 obtained from ovarian cancer tissue compared with normal ovarian tissue (Fig. 13A). The lower panel of Figure 15 is a streptavidin (SA-HRP) interacting protein for each lane. These results demonstrate that equivalent levels of protein were present in the AAL precipitation reactions. Therefore, the increased abundance of Lamp-1 in the AAL precipitation from serum of women with ovarian cancer is not due to increased protein input. These results
10 provide evidence that the fucosylation of N-linked glycans on Lamp-1 is an ovarian tumor-specific marker found in serum.

The complete disclosure of all patents, patent applications, and publications, and
15 electronically available material (including, for instance, nucleotide sequence submissions in, e.g., GenBank and RefSeq, and amino acid sequence submissions in, e.g., SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq) cited herein are incorporated by reference. In the event that any inconsistency exists between the disclosure of the present application and the disclosure(s) of any document incorporated
20 herein by reference, the disclosure of the present application shall govern. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

25 Unless otherwise indicated, all numbers expressing quantities of components, molecular weights, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless otherwise indicated to the contrary, the numerical parameters set forth in the specification and
30 claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the

doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques.

5 Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. All numerical values, however, inherently contain a range necessarily resulting from the standard deviation found in their respective testing measurements.

10 All headings are for the convenience of the reader and should not be used to limit the meaning of the text that follows the heading, unless so specified.

Sequence Listing Free Text

SEQ ID NOs: 1-307

peptide fragments

SEQ ID NOs: 308-347

oligonucleotide primer

WHAT IS CLAIMED IS:

1. A method for evaluating the presence, absence, nature or extent of cancer or a precancerous condition in a subject, the method comprising:
detecting the presence of a cancer-specific glycoform of a glycoprotein in a biological sample obtained from the subject, wherein the cancer-specific glycoform comprises a glycan indicative of the presence of cancer or a precancerous condition.
2. The method of claim 1 wherein detecting the cancer-specific glycoform of the glycoprotein comprises contacting the biological sample with a glycan-binding molecule specific for the glycan, under conditions that permit binding of the cancer-specific glycoform of the glycoprotein to the glycan-binding molecule.
3. The method of claim 1 wherein the cancer is selected from breast cancer, ovarian cancer, colorectal cancer, pancreatic cancer and liver cancer.
4. The method of claim 1 wherein the glycan-binding molecule is selected from the group consisting of a lectin, a glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, and a glycospecific small molecule.
5. The method of claim 4 wherein the cancer is breast cancer and the lectin comprises leucoagglutinating phytohemagglutinin (L-PHA).
6. The method of claim 5 wherein the glycan comprises a GlcNAc β (1,6) Man branched *N*-linked glycan.
7. The method of claim 5 wherein the glycan comprises a branched *N*-linked glycan extended with *N*-acetylactosamine.
8. The method of claim 1 wherein the cancer is breast cancer and wherein the glycoprotein comprises at least one glycoprotein selected from the group set forth in Table 4.

9. The method of claim 8 wherein the glycoprotein comprises at least one glycoprotein selected from the group consisting of periostin and osteoglycin.
10. The method of claim 4 wherein the cancer is ovarian cancer and the lectin comprises at least one lectin selected from the group consisting of erythroagglutinating phytohemagglutinin (E-PHA), Aleuria aurantia lectin (AAL) and Datura stramonium lectin (DSL).
11. The method of claim 10 wherein the glycan comprises an $\alpha(1,6)$ -fucose linked to core *N*-acetylglucosamine (core fucosylation).
12. The method of claim 10 wherein the glycan comprises a GlcNAc $\beta(1,4)$ Man bisected *N*-linked glycan.
13. The method of claim 1 wherein the cancer is ovarian cancer and wherein the glycoprotein comprises at least one glycoprotein selected from the group set forth in Table 8.
14. The method of claim 1 wherein the biological sample comprises a biological fluid.
15. The method of claim 1 wherein the biological sample comprises serum or plasma.
16. The method of claim 1 wherein the biological sample comprises tissue.
17. The method of claim 1 wherein the subject is a human.
18. A method for identifying a biomarker associated with cancer in a subject comprising:
 - providing a biological sample obtained from a subject, wherein the biological sample comprises a cancer-specific glycoform of a glycoprotein comprising a glycan; and
 - contacting the biological sample with a glycan-binding molecule specific for the glycan, under conditions that permit binding of the cancer-specific glycoform of the

glycoprotein to the glycan-binding molecule to yield a bound glycoprotein/glycan-binding molecule complex.

19. The method of claim 18 further comprising:

providing a noncancerous biological sample from the subject;

contacting the noncancerous biological sample with the glycan-binding molecule specific for the glycan, under conditions that permit binding of the cancer-specific glycoform of the glycoprotein to the glycan-binding molecule to yield a bound glycoprotein/glycan binding molecule complex; and

determining whether the amount of bound glycoprotein from the tumor sample is greater than the amount of bound glycoprotein from the noncancerous sample, wherein a greater amount in the tumor sample indicates the presence of a cancer-specific glycoform of the glycoprotein.

20. The method of claim 18 wherein the glycan-binding molecule is selected from the group consisting of a lectin, a glycospecific antibody, a glycospecific aptamer, a glycospecific peptide, and a glycospecific small molecule.

21. The method of claim 20 wherein the lectin comprises a lectin selected from the group consisting of L-PHA, E-PHA, AAL and DSL.

22. The method of claim 18 wherein the glycan is selected from the group consisting of a GlcNAc β (1,6) Man branched *N*-linked glycan component, a GlcNAc β (1,4) Man bisected *N*-linked glycan component, an α (1,6) fucose linked to a core *N*-acetylglucosamine, and a branched *N*-linked glycan extended with *N*-acetyllactosamine.

23. A cancer-specific glycoform of periostin comprising a GlcNAc β (1,6) Man branched *N*-linked glycan component.

24. A cancer-specific glycoform of osteoglycin comprising GlcNAc β (1,6) Man branched *N*-linked glycan component.

25. A cancer-specific glycoform of a glycoprotein set forth in Table 4, said glycoform at least one of a GlcNAc β (1,6) Man branched *N*-linked glycan component and a branched *N*-linked glycan extended with N-acetyllactosamine.

26. A cancer-specific glycoform of a glycoprotein set forth in Table 8, said glycoform comprising at least one structural feature selected from the group consisting of a GlcNAc β (1,4) Man bisected *N*-linked glycan component and an α (1,6) fucose linked to a core *N*-acetylglucosamine.

27. A multiplexed method for evaluating the presence, absence, nature or extent of cancer or a precancerous condition in a subject, the method comprising:

providing a biological sample obtained from the subject, the biological sample comprising a plurality of glycoproteins;

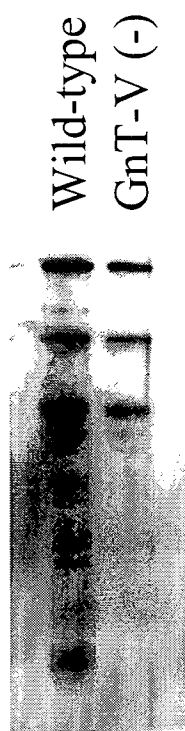
for each of the plurality of glycoproteins, determining the presence, absence or amount of a cancer-specific glycoform of the glycoprotein in the biological sample, wherein the presence, absence or amount of the cancer-specific glycoform is indicative of cancer or a precancerous condition.

28. The multiplexed method of claim 27 wherein the cancer is breast cancer, and wherein the plurality of glycoproteins comprises periostin and osteoglycin.

29. The multiplexed method of claim 27 wherein the cancer is ovarian cancer, and wherein the plurality of glycoproteins comprises at least two proteins, each independently selected from the proteins set forth in Table 8.

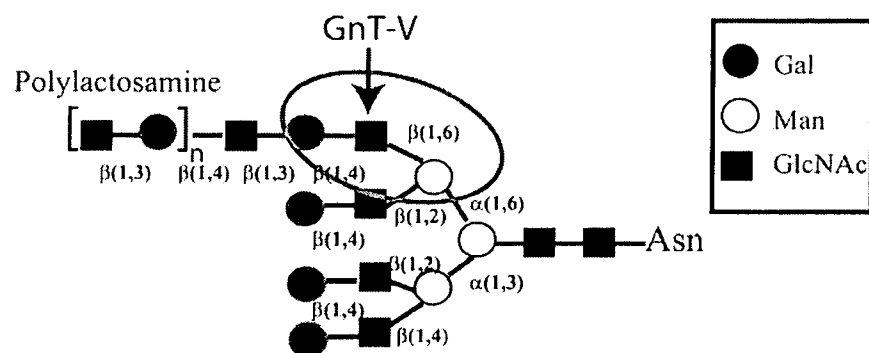
1/15

Fig. 1



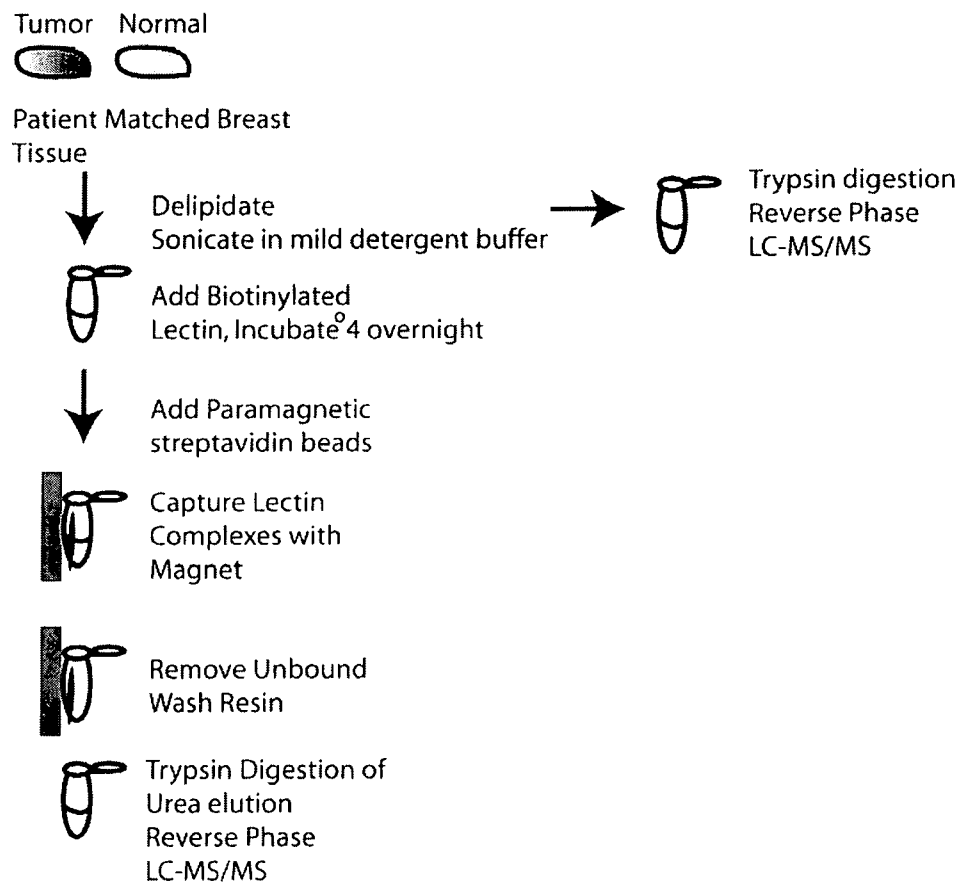
2/15

Fig. 2



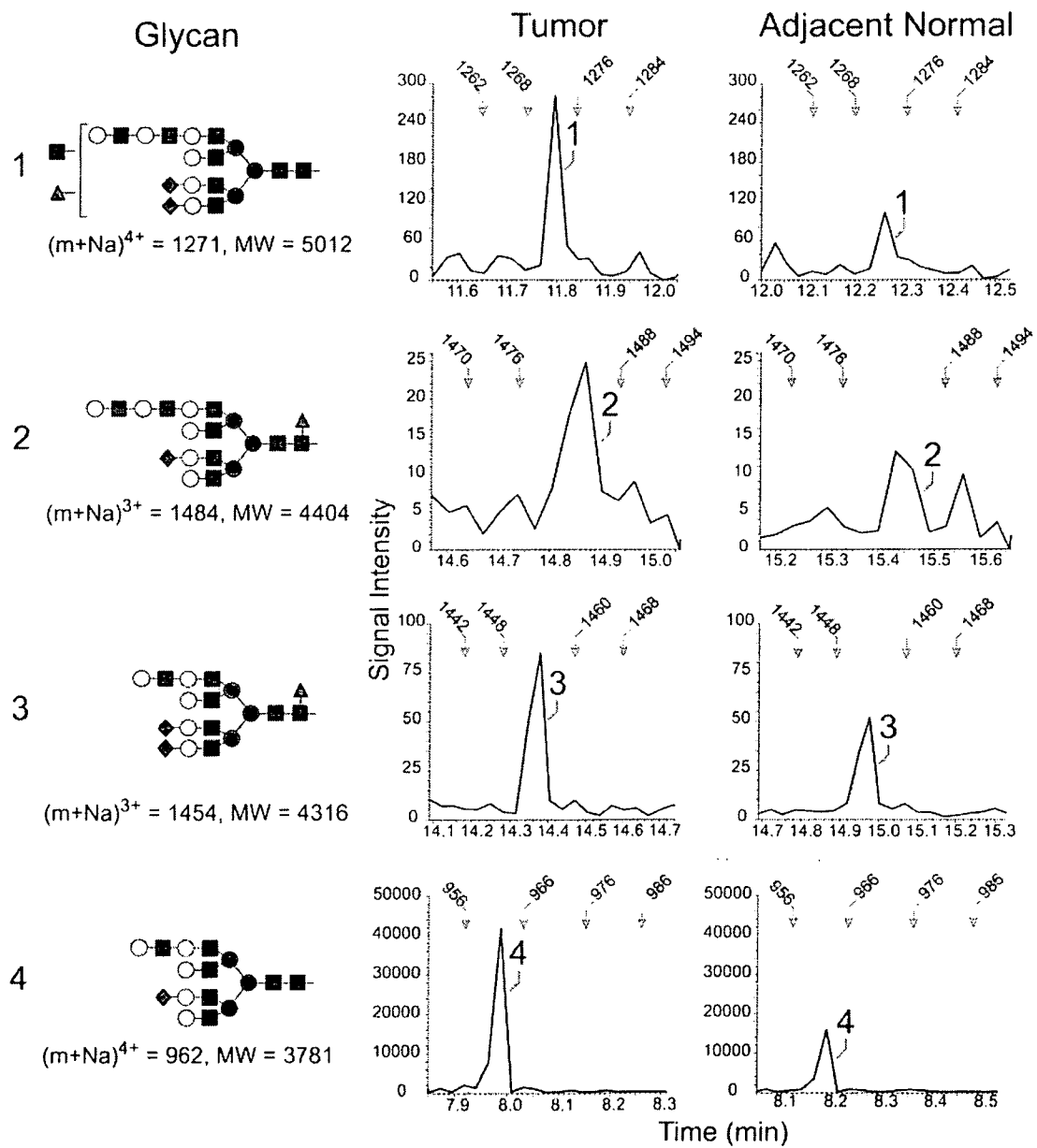
3/15

Fig. 3



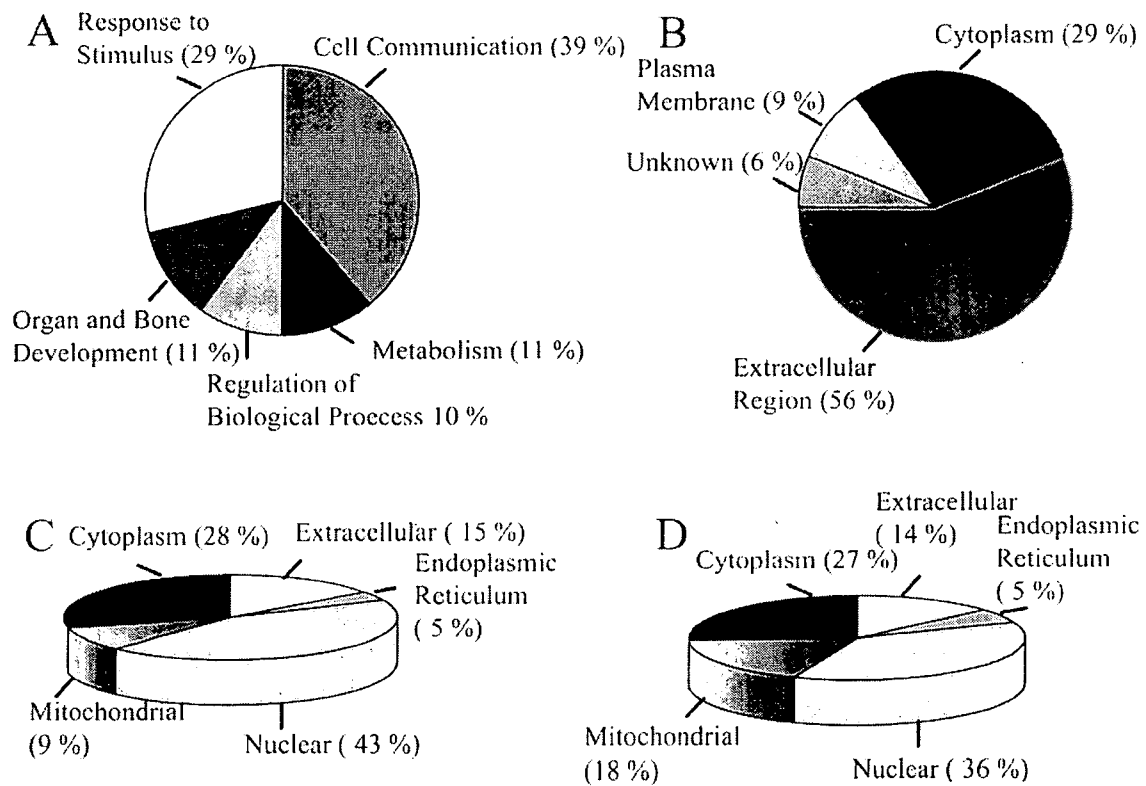
4/15

Fig. 4



5/15

Fig. 5



6/15

Fig. 6

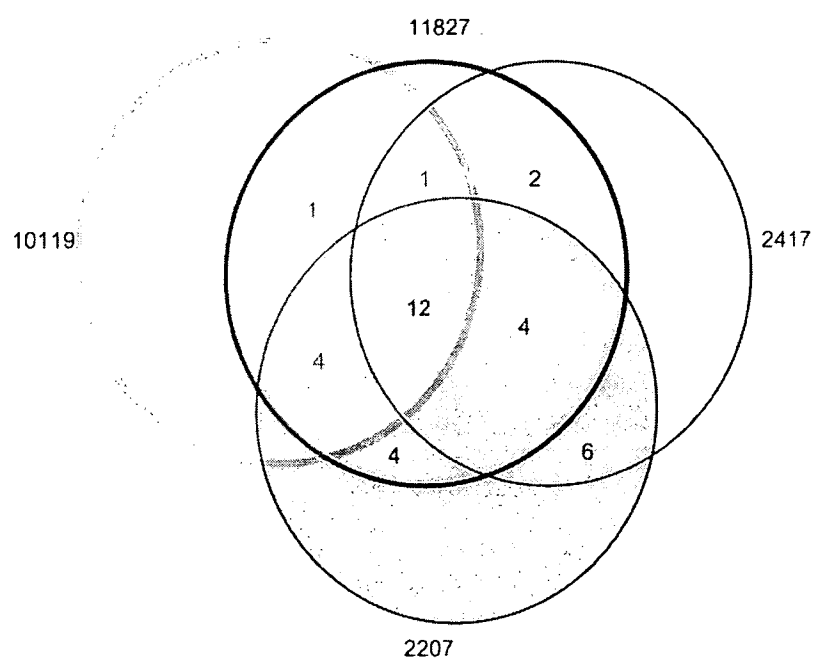
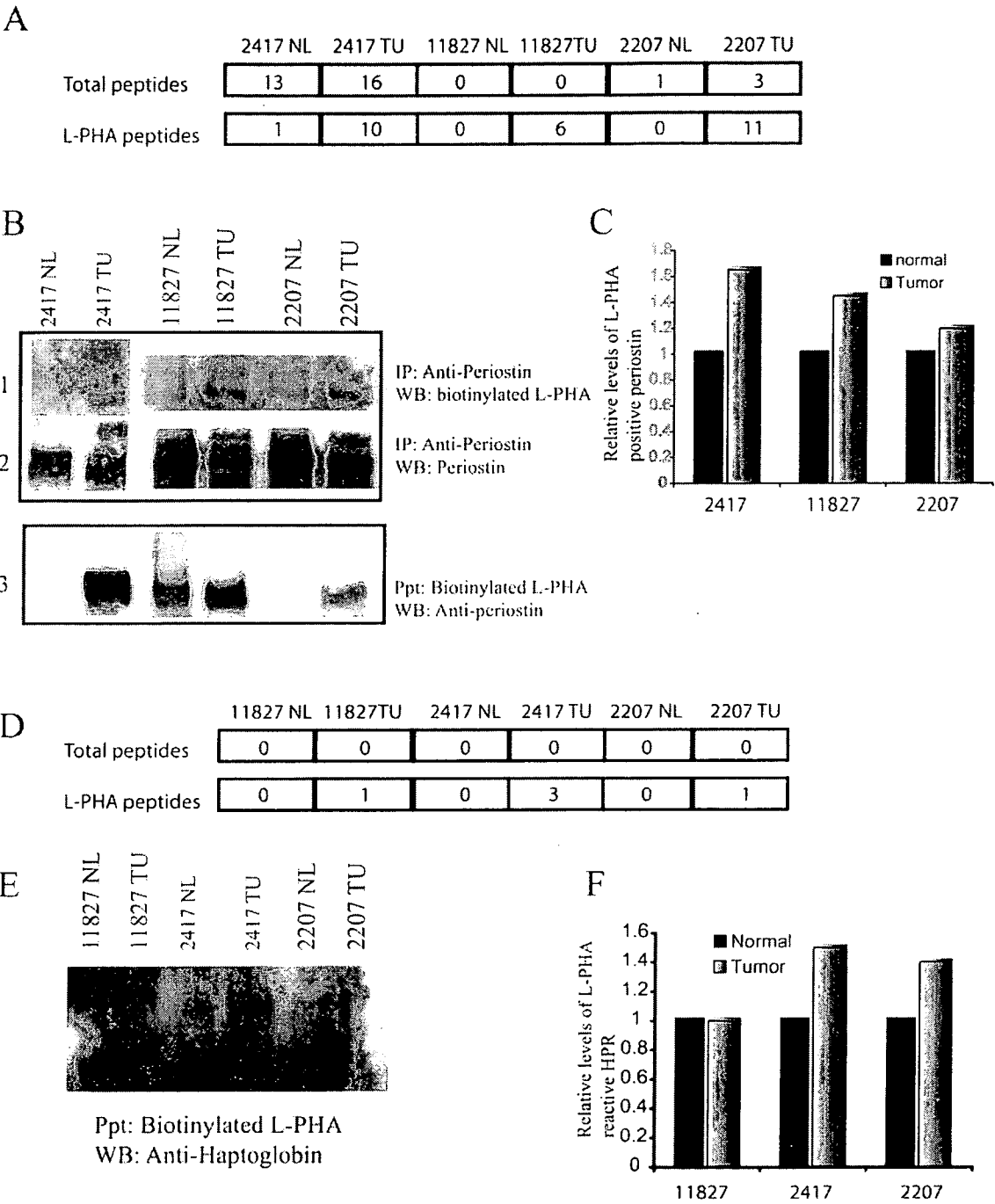
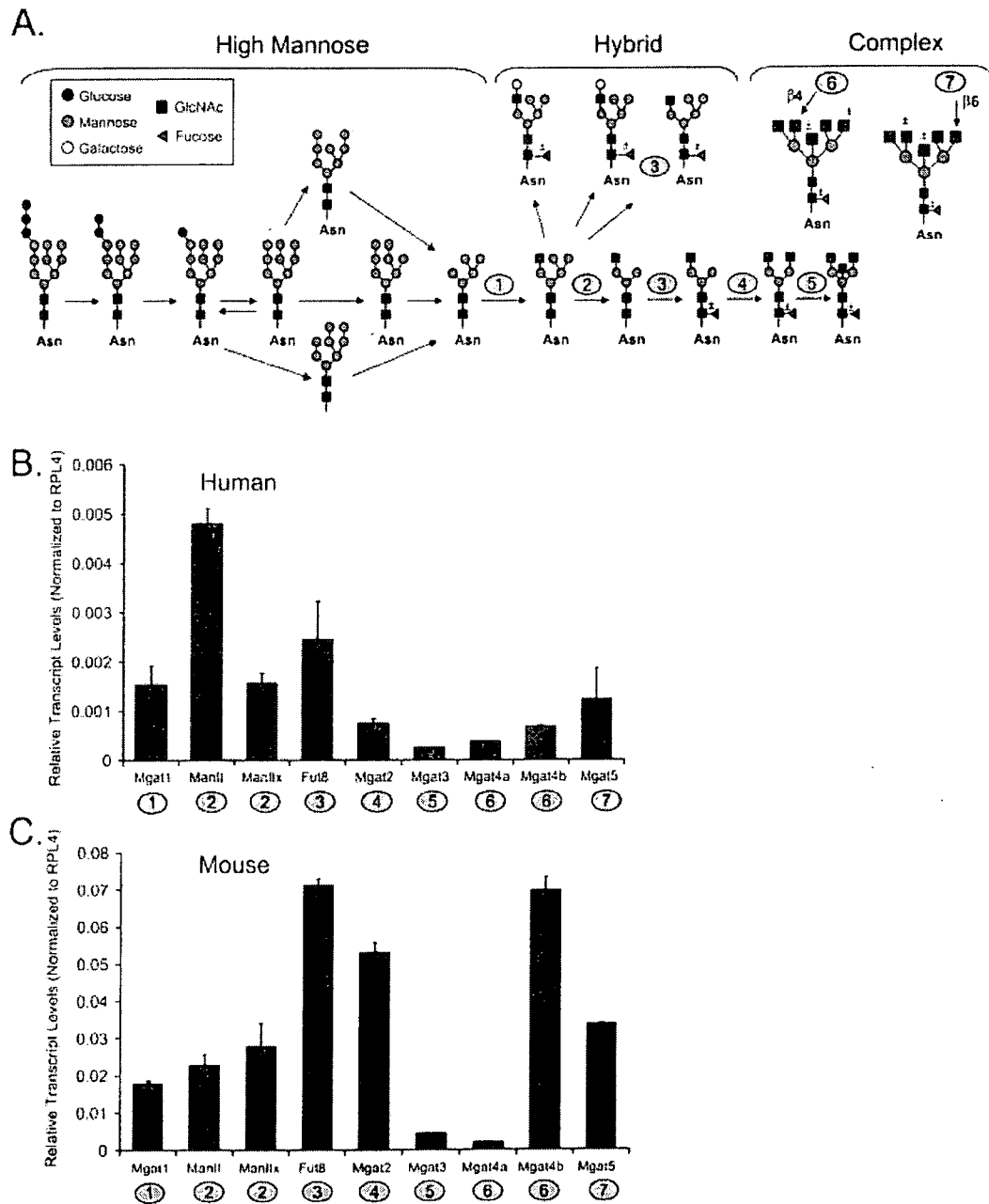


Fig. 7



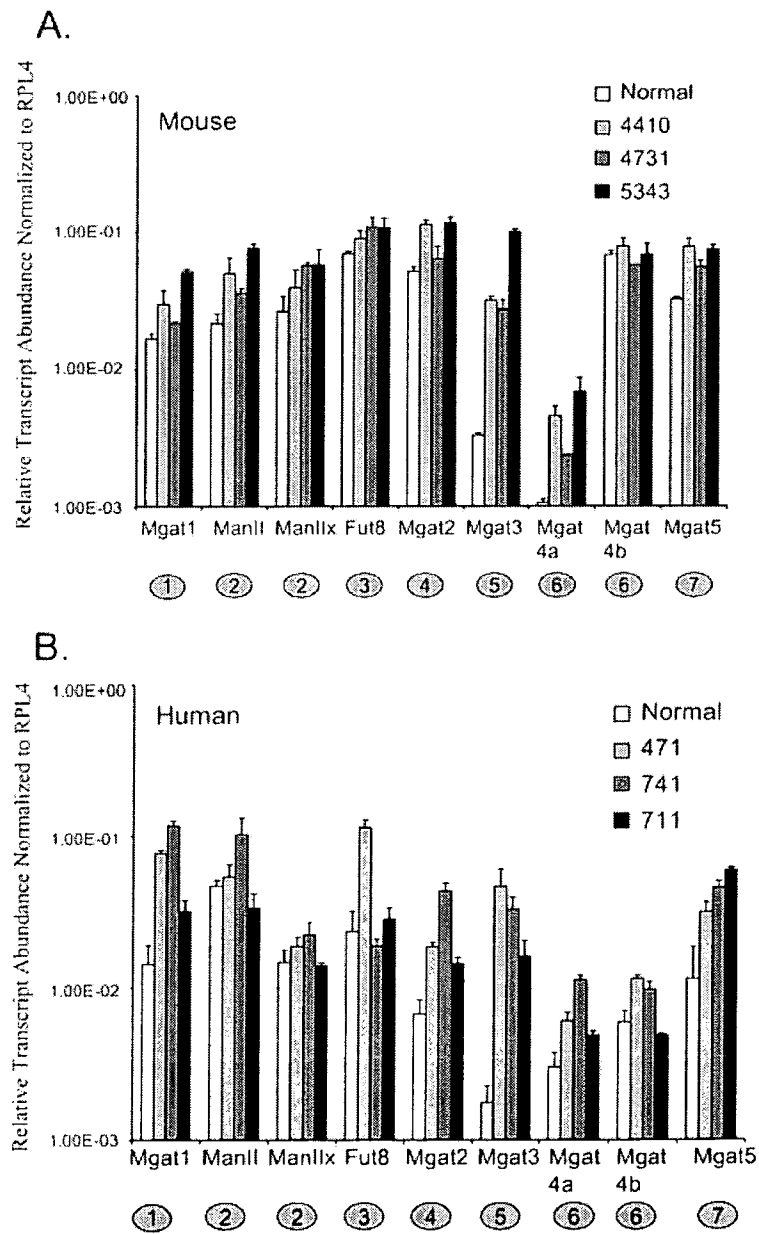
8/15

Fig. 8



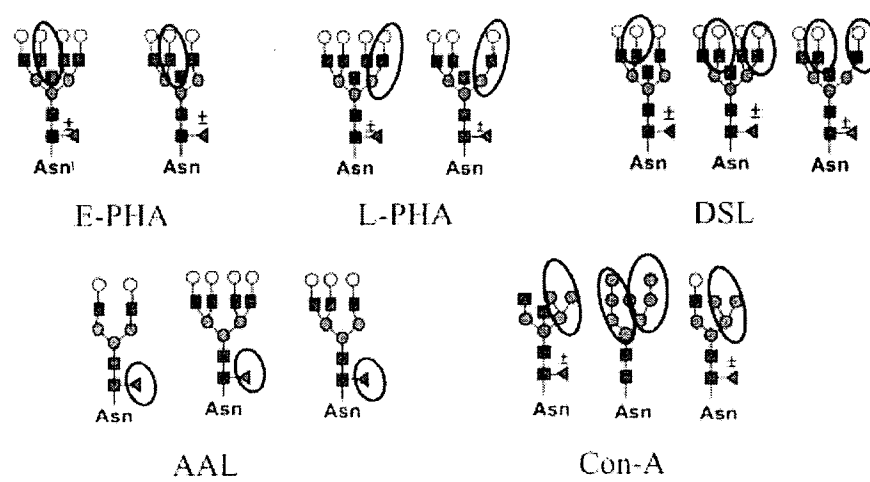
9/15

Fig. 9



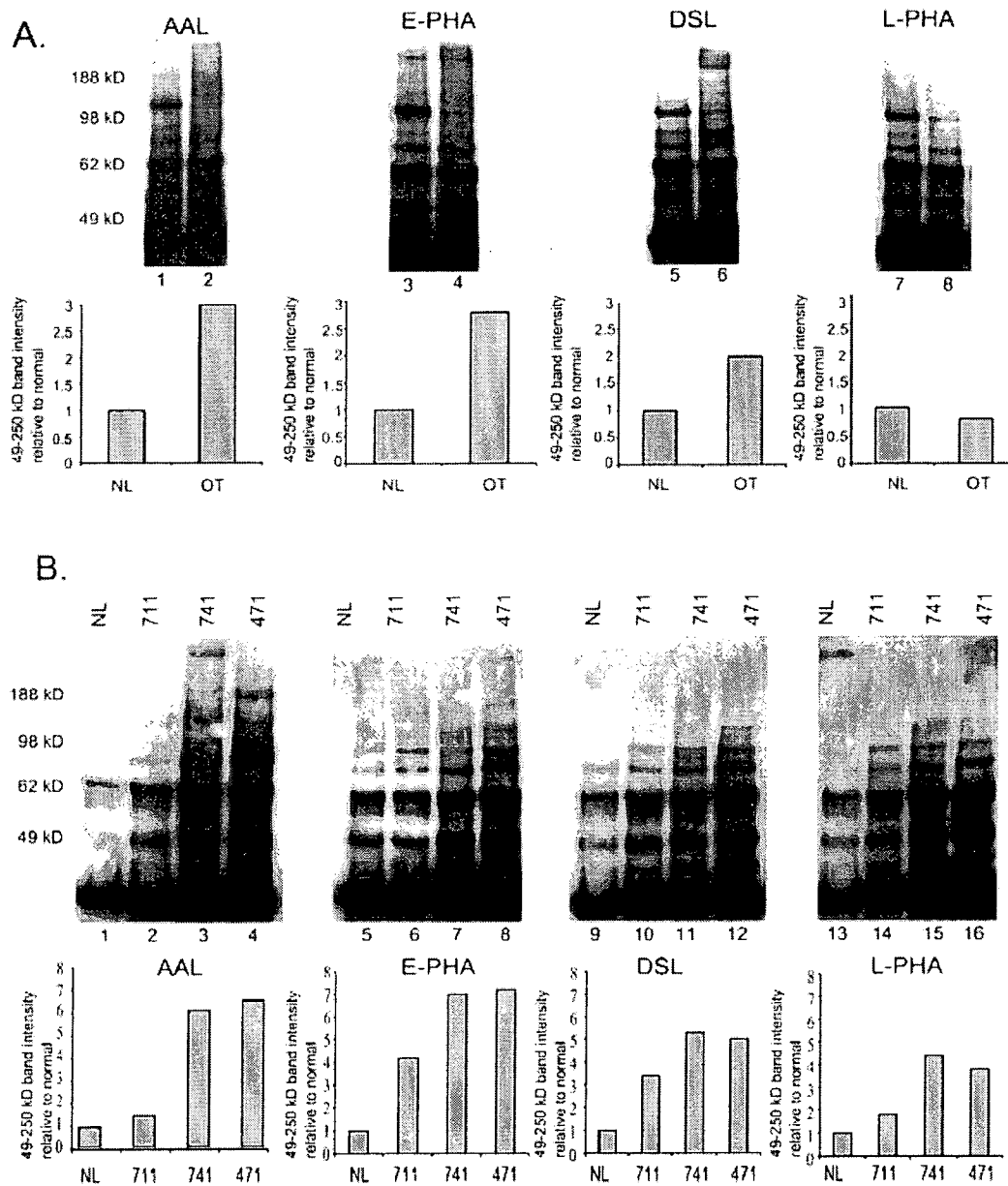
10/15

Fig. 10



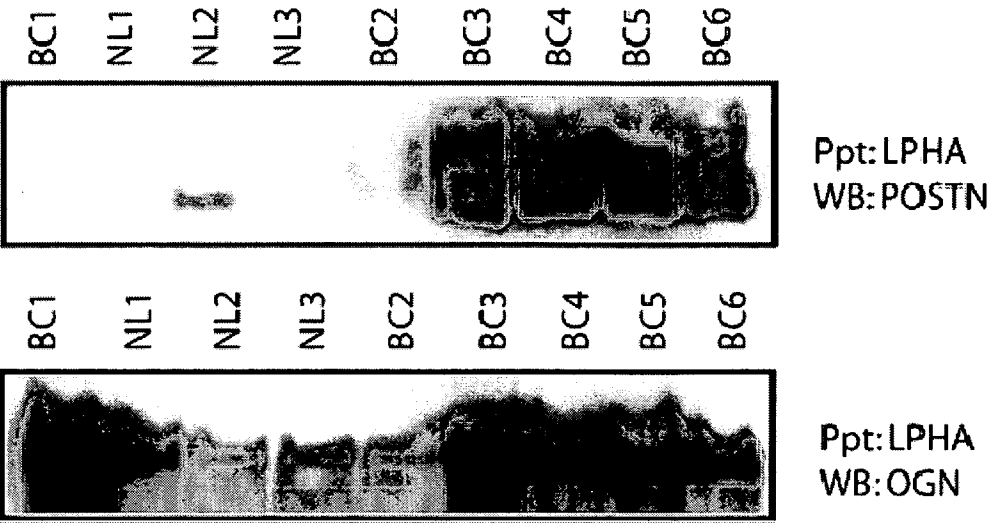
11/15

Fig. 11



12/15

Fig. 12



13/15

Fig. 13

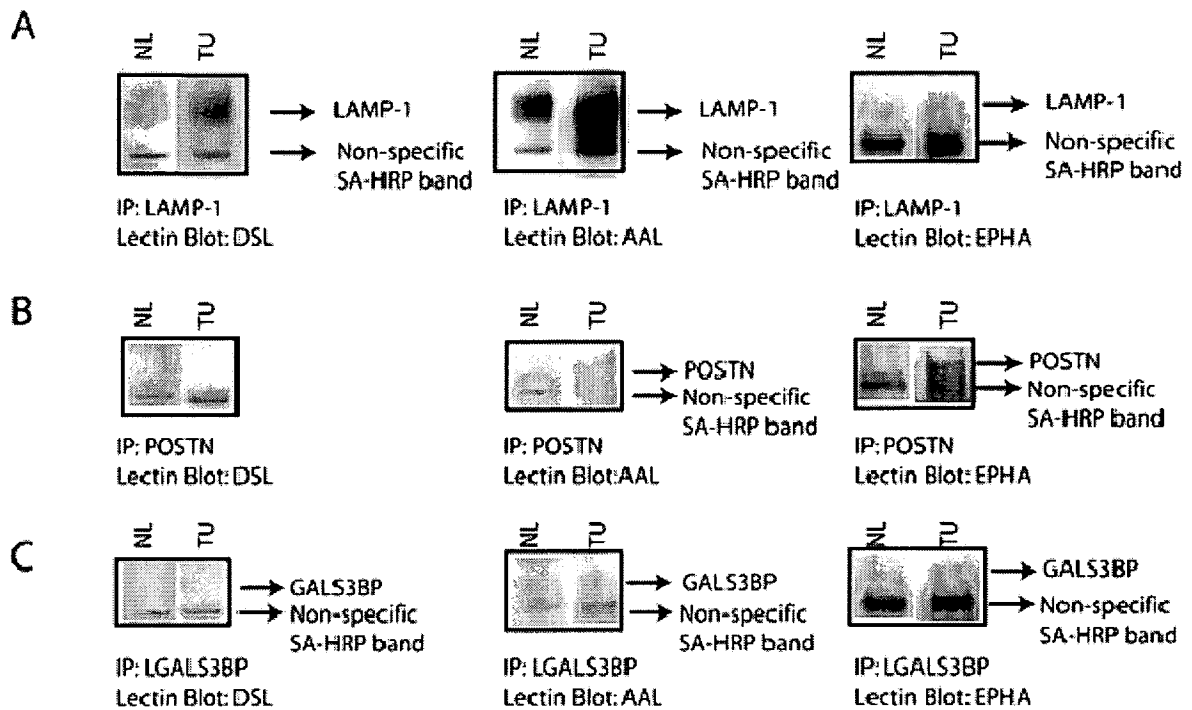
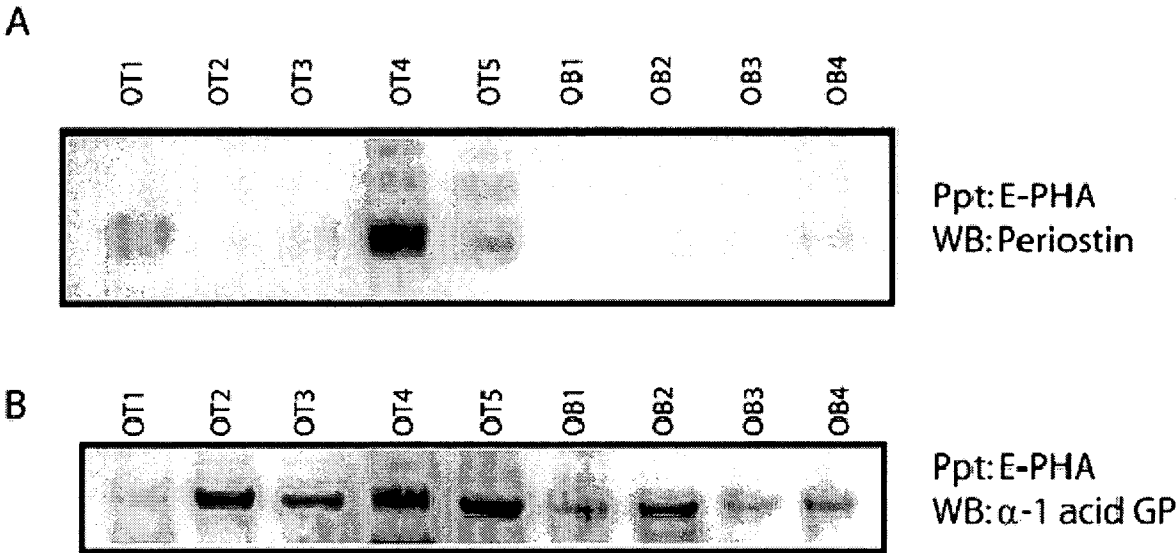


Fig. 14



15/15

Fig. 15

