Published in final edited form as:

J Proteome Res. 2012 April 6; 11(4): 2508–2520. doi:10.1021/pr201206w.

A lectin chromatography/mass spectrometry discovery workflow identifies putative biomarkers of aggressive breast cancers

Penelope M. Drake^{*}, Birgit Schilling[†], Richard K. Niles^{*}, Akraporn Prakobphol^{*}, Bensheng Li[†], Kwanyoung Jung[‡], Wonryeon Cho^{**}, Miles Braten^{*}, Halina D. Inerowicz[‡], Katherine Williams^{*}, Matthew Albertolle^{*}, Jason M. Held[†], Demetris Iacovides[§], Dylan J. Sorensen[†], Obi L. Griffith[§], Eric Johansen^{*}, Anna M. Zawadzka[†], Michael P. Cusack[†], Simon Allen^{*}, Matthew Gormley^{*}, Steven C. Hall^{*}, H. Ewa Witkowska^{*}, Joe W. Gray^{§,¶}, Fred Regnier[‡], Bradford W. Gibson^{†,||,††}, and Susan J. Fisher^{*,††}

^{*}Department of Obstetrics, Gynecology and Reproductive Sciences, 513 Parnassus Ave., Box 0665, University of California San Francisco, San Francisco, CA 94143

[†]Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, CA 94945

[‡]Department of Chemistry and Bindley Bioscience Center, 201 S. University St. HANS B054, Purdue University, West Lafayette, IN 47907

^{**}Bio-Nano Chemistry, Wonkwang University, 344-2 Shinyong-dong, Iksan, Jonbuk 570-749, Korea

[§]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[¶]Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97238

^{||}Department of Pharmaceutical Chemistry, Box 0446, University of California, San Francisco, CA 94143

Abstract

We used a lectin chromatography/MS-based approach to screen conditioned medium from a panel of luminal (less aggressive) and triple negative (more aggressive) breast cancer cell lines (n = 5/ subtype). The samples were fractionated using the lectins *Aleuria aurantia* (AAL) and *Sambucus nigra* agglutinin (SNA), which recognize fucose and sialic acid, respectively. The bound fractions were enzymatically N-deglycosylated and analyzed by LC-MS/MS. In total, we identified 533 glycoproteins, ~90% of which were components of the cell surface or extracellular matrix. We observed 1011 glycosites, 100 of which were solely detected in 3 triple negative lines. Statistical analyses suggested that a number of these glycosites were triple negative-specific and thus potential biomarkers for this tumor subtype. An analysis of RNAseq data revealed that approximately half of the mRNAs encoding the protein scaffolds that carried potential biomarker glycosites were upregulated in triple negative vs. luminal cell lines, and that a number of genes encoding fucosyl- or sialyltransferases were differentially expressed between the two subtypes, suggesting that alterations in glycosylation may also drive candidate identification. Notably, the glycoproteins from which these putative biomarker candidates were derived are involved in cancer-related processes. Thus, they may represent novel therapeutic targets for this aggressive tumor subtype.

^{††}To whom correspondence should be address: Susan Fisher, phone: (415) 476-5297, fax: 415-476-5623, sfisher@cgl.ucsf.edu; Bradford W. Gibson, phone: (415) 209-2032, fax: (415) 209-2231, bgibson@buckinstitute.org.

Introduction

The intense interest in biomarker discovery is a reflection of the clinical need for tests with a high degree of sensitivity and specificity for diagnosing diseases, predicting their courses, as well as monitoring responses to therapy and disease recurrence. Technological breakthroughs in separation strategies and mass spectrometry (MS) have enabled rapid identification and quantification of large numbers of proteins in biological samples¹. Nonetheless, their complexity requires extensive fractionation to access low abundance proteins, such as those released from nascent tumors. Alternatively, and technically less challenging, is the design of capture approaches that exploit disease biology for the purpose of biomarker identification². For many reasons, glycosylation is an attractive target. First, the biology allows for the rational design of discovery efforts. For example, changes in the glycosylation machinery can be identified from microarray data and translated in structural terms, providing a compelling rationale for designing lectin-based strategies to enrich glycopeptides carrying disease-related carbohydrate motifs. Second, one protein can carry many copies of an altered glycan, which may also be added to other scaffolds. Thus, there is an important amplification effect, which could enable the detection of many fewer abnormal cells than would otherwise be possible. Finally, glycosylation acts to shield the peptide backbone from proteolytic degradation³. Thus, in theory, glycan-based biomarkers are likely to be more stable in a variety of disease settings than unmodified proteins, which are often more labile.

Glycosylation is altered in a number of pathologies, but its relationship to cancer is particularly well-defined at phenotypic and, to a lesser degree, functional levels. For example, many of the most widely used clinical tests detect glycoproteins and carbohydrate structures. These include carcinoembryonic antigen (CEA), commonly used as a marker of colorectal cancer; CA-125, frequently employed to diagnose ovarian cancer; CA 19-9, the most commonly used biomarker for diagnosing pancreatic cancer; CA 15-3, used to monitor the metastasis of breast cancer⁴; and prostate-specific antigen (PSA)⁵⁻⁸. In addition, glycan-specific antibodies and lectins are used for the cytological and histological evaluation of glycosylation for the purpose of guiding diagnoses and enabling more accurate prognoses, e.g., anti-Lewis (Le)^x antibodies for bladder cancer, and the lectins *Helix pomatia* agglutinin (HPA) and *Ulex europaeus* I agglutinin (UEA 1) for breast cancer¹. This is due to the fact that increases in fucosylation and sialylation of *N*-linked structures and truncation of *O*-linked oligosaccharides occur in many tumor types. The expression of Le antigens, such as sialyl Le^x, can also be indicative of disease progression, as these structures play important roles in promoting metastasis by virtue of their well-known ability to mediate cell trafficking and extravasation^{9, 10}.

Breast cancer is now recognized to be a collection of distinct neoplastic diseases with different molecular and clinical attributes. Breast tumors can be stratified into five intrinsic subtypes and a “normal-like” group according to features such as mRNA expression¹¹. Interestingly, these molecularly-defined cohorts, which include luminal, basal-like, and claudin-low, are also predictive of clinical outcomes such as disease severity and treatment response¹²⁻¹⁴. Specifically, luminal tumors tend to be less aggressive with better survival rates, while basal-like and claudin-low lesions have generally worse prognoses¹⁵. Additionally, the expression of a therapeutic target such as the estrogen receptor (ER) or human epidermal growth factor receptor 2 (HER2/ErbB2) determines tumor susceptibility to drugs that interact with these molecules^{16, 17}. Triple negative breast cancers (TNBC) express neither ER nor the progesterone receptor (PR) and moderate levels of HER2. This clinically important, heterogeneous category includes most basal-like and claudin-low tumors^{18, 19}. TNBCs have poor survival rates and lack specific therapeutic targets, limiting treatment options and making early detection a priority.

We hypothesized that biomarkers specific for these tumors could be identified by a comparative analysis of the repertoire of secreted or shed glycoproteins in a panel of breast cancer cell lines that have been extensively characterized at genomic and transcriptional levels^{20–22}. Based on gene expression, the lines can be clustered into subsets that mirror the molecular characteristics of primary breast tumors. Thus, these panels are useful tools for studying subtype-specific behavior, such as drug responses and alternative splicing^{20, 23}. Here, we used a subset of cells from this collection for biomarker discovery. Specifically, we analyzed conditioned medium (CM) from 5 luminal and 5 triple negative cell lines. The samples were distributed to three laboratories: University of California San Francisco (UCSF), the Buck Institute for Research on Aging, and Purdue University. Each group analyzed the samples using our recently published method for lectin affinity chromatographic enrichment and LC-MS/MS analyses²⁴. Overall, we identified 533 glycoproteins, including 1011 *N*-linked glycosylation sites (glycosites). Of these, 100 were solely detected in 3 triple negative lines. Interestingly, many in the latter category were from glycoproteins that are upregulated in the claudin-low subtype²¹, involved in cancer progression (*e.g.*, epithelial to mesenchymal transition) and/or metastasis,²⁵.

Materials and Methods

Cell culture and production of conditioned media

All cells were cultured as described in Neve *et al.*²¹. To generate the CM, we cultured 10 breast cancer cell lines (Table 1) that were derived from 5 luminal (SKBR3, SUM52 PE, MDAMB175, UACC 812, and MDAMA361) and 5 triple negative tumors (MDA468, BT549, HS578T, MDAMB231, and HCC38). CM was prepared and trypsin digested at Site M. The lines were grown to 75–80% confluence in the appropriate culture medium²¹. Then they were washed with fresh medium without fetal calf serum (FCS) or phenol red and incubated for 10 min at 37 °C. This process was repeated twice before the cells were incubated in fresh medium (without FCS and phenol red) for 18–20 h. At the end of the culture period, the cells retained their original morphologies with no evidence of apoptosis. The CM was harvested and centrifuged at 2000 × *g* for 10 min. The supernatant was concentrated using Millipore centrifugal filter units (MWCO 3K) and dialyzed against phosphate buffered saline (PBS).

Lectin blotting and staining

Biotinylated and fluoresceinated lectins were purchased from Vector Laboratories. Blotting: Cell lysates were separated by SDS-PAGE (4–12% gels) and transferred to nitrocellulose membranes. Unless otherwise indicated, the following buffer was used for all steps, including blocking, washing, and reagent dilution/incubation: 0.25 M Tris-Cl, pH 8.0, 0.5 M NaCl, 0.5% NP-40. Blots were incubated in buffer for 1 h to block non-specific binding, then exposed to a solution of ~5 µg/mL of biotinylated lectin for 2 h. Blots were washed 3 × 5 min with copious amounts of buffer. Then, membranes were reacted with ABC reagent (Vector Laboratories) for 1 h and washed again as before. Finally, bound lectin was detected using 3,3-diaminobenzidine (DAB, Vector Laboratories) prepared in water according to the manufacturer's instructions. Staining: cell surface labeling of non-permeabilized cells was performed as described²⁶, except that fluoresceinated lectins, rather than antibodies, were used.

Trypsin digestion

First, protein concentrations of the CM samples were determined by amino acid analysis. Then, CM samples were digested and desalted using a published method that incorporates denaturation with 6 M urea²⁷. As previously described²⁴, samples were spiked with 25 and 50 pmol of trypsin-digested control glycopeptides from commercial yeast invertase and

human lactoferrin (Sigma, St. Louis, MO), respectively. Peptides were stored at -80°C prior to analyses.

Preparation of lectin columns

The columns were prepared at Site M from a single batch of lectin-conjugated beads and distributed to all the laboratories. Briefly, *Sambucus nigra* agglutinin (SNA) and *Aleuria aurantia* lectin (AAL) were purchased from Vector Laboratories (Burlingame, CA). Lectins (20 mg) were suspended at 5–10 mg/mL in PBS and conjugated to 330 mg of POROS-AL beads (Applied Biosystems, Foster City, CA) as previously described²⁴. Unconjugated protein was removed by washing the beads (5×5 mL of 1 M sodium chloride) before they were packed into 3 individual 4.6×50 mm PEEK HPLC columns. Routine storage was in PBS with 0.02% sodium azide at 4°C for up to 6 months. Columns were reused for up to 75 affinity separations without degradation of the performance characteristics as assessed by glycopeptide enrichment and total number of glycopeptides recovered from digested human plasma.

Lectin chromatography—Instrumentation

The HPLC systems employed were standardized in terms of injection volume, transfer line lengths, dead volume minimization, and common UV elution profiles. Site M used a Paradigm MG4 HPLC system equipped with a CTC PAL robot configured as an autosampler and fraction collector (Michrom Bioresources). At Site X, a Waters system including 1525 Binary HPLC equipped with a 717 plus Autosampler and a Fraction Collector III was employed. Site S used a Shimadzu 20AD HPLC system equipped with a SIL-20AC autosampler; fractions were collected manually. *Mobile phases*: Buffer A was 25 mM Tris buffer, pH 7.4, 50 mM sodium chloride, 10 mM calcium chloride, and 10 mM magnesium chloride; Buffer B was 0.5 M acetic acid. *Affinity separation*: Routinely, ~ 100 μg of digested protein was diluted into Buffer A, applied to the lectin column, and separated using the following 3 step gradient: 1) Sample load: Buffer A for 9.0 min at 80 $\mu\text{L}/\text{min}$; 2) Sample elution: Buffer B for 4.8 min at 500 $\mu\text{L}/\text{min}$; and 3) Re-equilibration: Buffer A for 6.0 min at 3000 $\mu\text{L}/\text{min}$. The bound fraction, collected from 9.0 to 14.25 min, was desalted using Oasis HLB cartridges as described above. Eluted samples were neutralized by the addition of 0.5 M ammonium bicarbonate and concentrated to <100 μL by vacuum centrifugation. Further details are described in the accompanying SOP (Supplementary Document 1).

PNGase F digestion

N-linked glycopeptides in the bound fractions were deglycosylated by treatment with PNGase F (Glycerol-free, New England Biolabs; Ipswich, MA) as previously described²⁴. Following deglycosylation, samples were desalted and concentrated using C18 ZipTips® (Millipore; Billerica, MA) or MicroSpin Columns, 5–200 μL (The Nest Group, Inc.; Southborough, MA).

ESI-QqTOF mass spectrometric analyses (Sites M and X)

The peptides were separated using an Eksigent nano-LC 2D HPLC system (Eksigent, Dublin, CA), which was directly connected to a quadrupole time-of-flight (QqTOF) QSTAR Elite mass spectrometer (AB Sciex, Foster City, CA). We injected 33% (vol/vol) of the bound material per run. Briefly, peptides were applied to a guard column (C18 Acclaim PepMap100, 300 μm I.D. $\times$ 5 mm, 5 μm particle size, 100 Å pore size; Dionex, Sunnyvale, CA) and washed with the aqueous loading solvent (2% solvent B in A, flow rate: 20 $\mu\text{L}/\text{min}$) for 10 min prior to separation on a C18 Acclaim PepMap100 column (75 μm I.D. $\times$ 15 cm, 3 μm particle size, 100 Å pore size; Dionex, Sunnyvale, CA). Bound material was

eluted at a flow rate of 300 nL/min using the following gradients: 2–40% solvent B in A (from 0–60 min), 40–90% solvent B in A (from 60–75 min), and at 90% solvent B in A (from 75–85 min), with a total runtime of 120 min (including column equilibration). Solvent A consisted of 0.1% formic acid in 98% H₂O/2% acetonitrile and solvent B was 0.1% formic acid in 98% acetonitrile/2% H₂O. Spectra were calibrated using MS/MS fragmentations of a Glu-Fibrinogen B peptide standard. Advanced information dependent acquisition was employed for MS/MS data collection using QSTAR Elite (Analyst QS 2.0) specific features, including “Smart Collision” (fragment intensity multiplier set to 2.0) and “Smart Exit” (maximum accumulation time of 2.5 sec) to obtain MS/MS spectra for the six most abundant precursor ions following each survey scan. To increase overall sampling efficiencies, two replicate nano-HPLC-MS/MS analyses were performed per sample.

ESI-LTQ-Orbitrap XL mass spectrometric analyses (Site S)

The peptide mixtures were separated as described above using an Agilent nanoflow 1100 HPLC system (Agilent, Santa Clara, CA) connected to a hybrid linear ion trap Orbitrap mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific). The electrospray ionization emitter tip (Pico-tip emitter, F360-75-15-N-5-C10.5) was purchased from New Objective (Woburn, MA). The mass spectrometer, which was calibrated with a solution of caffeine, MRFA and Ultramark 1621 according to the manufacturer’s instructions, was operated in the data-dependent mode. Full MS scans from m/z 350 to 1600 with a full width at half maximum resolution of 30,000 were acquired as profile data, followed by MS/MS scans of the six most abundant ions in the linear trap. Singly charged ions were excluded. A dynamic mass exclusion time was applied for 120s with a repeat count of 1 and a repeat duration time of 30s. In all scan modes, one micro scan was applied.

Database searches

Mass spectrometric data from all laboratories were analyzed at Site M using two bioinformatics database search engines with integrated peak picking, ProteinPilot™ (AB Sciex) version 4.0.8085 (revision 148085) using the Paragon Algorithm 4.0.0.0, 148083²⁸, and Mascot version 2.2.04 using Mascot Daemon version 2.2.2 (both Matrix Science). For the latter, the following (default) data import filter options were used: precursor charge state +2 to +4, reject spectrum if < 7 peaks or if precursor is < 400 or >10000 m/z , remove peaks with intensity < 0.001% of the highest peak; centroid all MS/MS data, percentage height 50, and merge distance 0.1 atomic mass units. Peak lists for the Orbitrap LC-MS/MS data sets were generated using Mascot Distiller 2.3.2.0 (Matrix Science) with the supplied processing parameter file Orbitrap_low_res_MS2_4.opt. The Orbitrap peak lists were saved in MGF format with Distiller preferences set to save MS/MS peaks as MH+ for input into Mascot and ProteinPilot search engines. All data were searched using a merged database of 20293 protein sequences including the publicly available human SwissProt UniProt release 2010_09 plus 7 other proteins, which includes all 20,286 reviewed (formerly SwissProt) Human Uniprot Entries, as well as PNGase F (Q9XBM8|Q9XBM8_FLAME, P21163|PNGF_ELIMR) and Yeast Invertase (P10594|INV1_YEAST, P00724|INV2_YEAST, P10595|INV3_YEAST, P10596|INV4_YEAST, P10597|INV5_YEAST). *ProteinPilot* searches were performed as previously described²⁴. A ProteinPilot peptide confidence cut-off value of 98.8 was chosen, corresponding to a local FDR of 5%. For *Mascot* searches, the following parameters were used: trypsin enzyme specificity, carbamidomethyl (Cys) as a fixed modification, and the following variable modifications: deamidation of asparagine and glutamine residues, oxidation of methionines, acetylation at the protein *N*-terminus, cyclization of *N*-terminal glutamines, and two missed tryptic cleavages. For QSTAR Elite data a mass tolerance of 100 ppm and 0.4 Da was set for the precursor and product ions, respectively; whereas values of 10 ppm and 0.8 Da were applied to Orbitrap data. Peptide-spectral matches with expectation values <0.026 were accepted. FDR analysis was

performed using the Mascot automatic decoy search. In all cases, the peptide false-positive identification rate was <3%.

Glycopeptide assignment

Deglycosylated peptides were identified as previously described²⁴, on the basis of several criteria including the motif NxS/T, x = proline, in which Asn was converted to Asp (reported by the search engine as Asn deamidation), and the presence of at least one fragment ion encompassing the glycosite. To ensure inclusion of glycosites containing Lys and/or Arg in the X position (*e.g.*, NKT), which were likely to have been cleaved by trypsin, the amino-acid residue following the carboxy-terminal cleavage site was also considered. Peptides containing the motif NGS or NGT were excluded due to the fact that asparagine residues in that sequence are prone to chemical deamidation during overnight trypsin digestion²⁹. For all deglycosylated peptides the corresponding MS/MS spectra were manually examined using an adaptation of previously published criteria to ensure correct assignment^{24, 30}.

Generation of a list of candidate triple negative-specific glycosites

The selection criteria for triple negative-specific glycosites were subjected to a resampling, non-parametric statistical test in which no knowledge about the data's distribution is necessary, *e.g.*, the "bootstrap" technique³¹. The basic premise of this approach is to consider the null hypothesis that there is statistically no difference between the luminal and triple negative data sets, *e.g.*, that the two are random selections from the same population. To determine the expected FDRs, we applied 20,000 random permutations to the form:

Criterion n-m: *A glycosite satisfies criterion n-m if it is identified in n Luminal cell lines and in m TN cell lines.*

The results are shown in Supplementary Table 3.

Spectral Viewer: Skyline Spectral Library

An interactive Skyline spectral library file that contains all MS/MS spectra of deglycosylated peptides identified in this study been submitted as Supplemental Material. Skyline is an open source program³² available for free download at <http://proteome.gs.washington.edu/software/skyline>.

Exon expression array and RNAseq experiments

Whole transcriptome shotgun sequencing (RNAseq) was completed on nine of ten breast cancer cell lines (BT549, HCC38, HS578T, MDAMB231, MDAMB175VII, MDAMB361, SKBR3, SUM52PE and UACC812). Expression analysis was performed with the ALEXA-seq software package as previously described³³. On a per sample basis, an average of 58.7 million (76bp paired-end) reads passed quality control, and 37.6 million mapped to the transcriptome, which resulted in coverage of 40x across all known genes. Log₂ transformed estimates of gene-level expression were extracted for fucosyl- and sialyltransferase genes, and triple negative candidate biomarker targets that emerged from the N-glycosite workflow. Corresponding values indicating whether expression of a transcript was detected above background were also extracted. A 2-sided Student's *t*-test was used to compare log₂ transformed gene expression levels between the five luminal and the four triple negative cell lines. This comparison generated raw *p*-values, which were then adjusted for multiple comparisons using the Benjamini-Hochberg method for controlling FDRs³⁴. The adjustment was achieved with the `p.adjust(pvals,"fdr")` function in R version 2.12.1 (2010-12-16). Adjusted FDR *p*-values lower than 10% (0.1) were considered significant.

Results and Discussion

Workflow

These experiments utilized a lectin chromatography, MS-based approach that we recently optimized and published to identify candidate cancer biomarkers²⁴. Initially, we probed nitrocellulose transfers of electrophoretically-separated cell lysates of breast cancer lines established from triple negative and luminal tumor subtypes with a panel of nine lectins (SNA, AAL, *Vicia villosa*, *Phaseolus vulgaris* leucoagglutinating and erythroagglutinating, *Galanthus nivalis*, *Euonymus europaeus*, *Lycopersicon esculentum*, and *Arachis hypogaea*) that recognized either internal saccharide motifs or terminal sugars. The results showed that SNA (Fig. 1a) and AAL (data not shown), which bind motifs with sialic acid and fucose, respectively, reacted with a wide array of glycoproteins. Additionally, some glycoforms were enriched in lines that were derived from the tumors of the same subtype. Staining of intact non-permeabilized cells with fluorescein-conjugated SNA revealed strong surface labeling (Fig. 1b). Together, these results suggested that the breast cancer cell lines produced a large repertoire of glycoproteins that reacted with SNA or AAL, including cell-surface molecules poised to be shed or released.

Next, we used this workflow to compare CM samples from 5 luminal and 5 triple negative breast cancer cell lines to identify subtype-specific glycosites. The cells, listed in Table 1, are members of a well-annotated collection that have been used to define the gene expression profiles, drug sensitivities, and protein splicing patterns of the tumor types from which they were derived^{20, 21, 23}. Contrary to many other lectin-based approaches, the affinity capture step was performed at the glycopeptide, rather than the protein level, which decreased non-specific binding due to hydrophobic interactions, a phenomenon that we previously observed between lectins and intact proteins. Thus, the samples were trypsin-digested prior to HPLC separation on lectin-conjugated POROS. Then, the bound fraction was treated with peptide *N*-glycosidase F (PNGase F) to remove *N*-linked glycans prior to LC-MS/MS analyses. The results were analyzed using two search engines, ProteinPilot and Mascot, to identify peptides and their corresponding proteins²⁸. *N*-glycosylates were identified as described in the methods²⁹. Finally, each MS/MS spectrum was manually inspected for the presence of at least one fragment ion that encompassed an *N*-glycosylation site. Thus, this method identified the glycosite that carries an oligosaccharide with a lectin-binding motif and the corresponding protein. These rigorous criteria were key to making this method highly reproducible²⁴.

We know from our participation in the Clinical Proteomic Technologies for Cancer (CPTAC) network that analysis of the same sample at multiple sites on different platforms is one way to maximize identifications and test the robustness of a workflow^{35, 36}. The experimental strategy we used, which exploited this observation, is depicted in Fig. 2. CM samples were trypsin-digested and aliquoted at a single site (Fig. 2A). Lectin enrichment and LC-MS/MS analyses were carried out according to a Standard Operating Procedure (SOP, Supplemental Document 1) at each of three locations—University of California San Francisco, Buck Institute for Research on Aging, and Purdue University (Fig. 2B). Prior to initiating the study, each group evaluated the lectin capture step using a National Institute of Standards and Technology (NIST) human pooled plasma sample, which we have extensively characterized with respect to the SNA and AAL chromatographic profiles and the glycosite composition of the bound fractions²⁴. MS analyses yielded glycosite identifications and percent enrichment values (total glycopeptides/total peptides) within the expected range²⁴.

Two groups, M and X, acquired data using a QSTAR Elite QqTOF (AB Sciex), while the third, S, used an LTQ-Orbitrap (Thermo Fisher Scientific). The datasets were submitted to

Site M, where all the searches and bioinformatic analyses were completed (Fig. 2C). As the work progressed, two changes to the protocol were implemented. First, due to technical problems encountered during the initial analysis, a second preparation of CM samples was analyzed at two of the three locations (M and S). Second, sites M and S replaced ZipTips® with spin-cartridges for the desalting step that followed PNGase F digestion. This change was made in response to the fact that, in initial experiments, Site S routinely identified significantly more glycosites using this desalting method. All peptides and glycopeptides observed in these experiments are presented as supplemental data (Supplementary Table 1).

Identification of >500 cell-surface or secreted glycoproteins

We tabulated the MS identifications according to the CM samples in which they were detected. Summaries of the data, including the number of glycoproteins, glycopeptides and *N*-glycosites observed in each CM sample, and the percent glycopeptide enrichment, are shown in Figs. 3 and 4, and in Supplementary Table 2. Overall the three groups identified a total of 1011 distinct *N*-glycosites from 533 glycoproteins. Of these, 945 and 641 were observed following AAL and SNA chromatography, respectively. Interestingly, the same workflow applied to pooled healthy human plasma resulted in many fewer identifications. Approximately half the species captured from CM bound to both lectins; the remainder preferentially interacted either with AAL or SNA. (Fig. 3A). A similar phenomenon was observed when the *N*-glycosites were grouped according to tumor subtype (Fig. 3B and C). Thus, it was clear that employing multiple lectins in our workflow resulted in a greater number of identifications. Furthermore, the data showed that the luminal and triple negative samples contained substantially different lectin-reactive species.

An overall comparison of the data obtained for luminal and triple negative samples across the three sites showed relatively high levels of enrichment in both cases (Fig. 4). Importantly, very few intracellular proteins were identified, additional evidence that the cells were not undergoing apoptosis. Approximately 90% of the glycoproteins observed reside either at the cell surface (59%) or in the extracellular matrix (29%), suggesting that our strategy of using CM as a source of secreted and/or shed glycoproteins was successful (Fig. 5). Since we wanted to identify candidate cancer biomarkers, we were interested to find that a number of the identified species have functions that are relevant to tumor biology. For example, we observed proteinases, including cathepsins and ADAM family members; adhesion molecules, including cadherins and integrins; extracellular matrix components, including decorin and SPARC; and cytokines, including leukemia inhibitory factor and vascular endothelial growth factor C. Furthermore, some of the glycoproteins had been previously identified as putative breast cancer biomarkers, including CD44, galectin-3 binding protein, insulin-like growth factor binding protein 3, and tissue inhibitor of metalloproteinase 1³⁷⁻³⁹. We also identified clinically useful markers, such as HER2/ ErbB2, and the CA-125 antigen, MUC16, which is commonly used to screen for ovarian cancer, but can be also be upregulated in breast tumors^{40, 41}.

Identification of putative glycosite biomarkers of triple negative breast cancers

Next, we used statistical analyses to generate a list of putative triple negative-specific glycosites. Specifically, we performed a statistical analysis using resampling methods that tested 20,000 random permutations of the data. This process generated a table (Supplementary Table 3) with the number of “triple negative-specific” glycosites expected at random for any given set of selection criteria (*e.g.*, observed in “ 1 triple negative and 0 luminal” or “ 4 triple negative and 1 luminal”). This analysis allowed us to select parameters that maximized the identification of putative triple negative specific glycosites while controlling the FDR. In this context, we required that a glycosite be identified at least once in CM samples from 3 triple negative cell lines and not observed in luminal CMs.

Using these criteria, the computed FDR for both lectin capture steps was ~15%. This yielded 49 candidates that bound to SNA and 76 that bound to AAL (Fig. 6). Of these, we removed glycosites from highly polymorphic HLA class I histocompatibility antigens, which are variably expressed in the population. The final list of 100 glycosites, from 83 glycoproteins, that were putative triple negative-specific candidates is shown in Table 2.

Next, we asked whether the glycosites we identified could have been predicted from transcriptome analyses. To answer this question, we used existing exon expression array profiles for all of the cell lines and RNAseq data for 9 of the 10. Since the two platforms identified similar sets of differentially expressed genes, we performed statistical analyses using values from the RNAseq experiments, which are better able to differentiate signal from noise (Supplementary Table 4). These analyses showed that 46 of the 83 mRNAs encoding the protein scaffolds that carried biomarker glycosites were upregulated 2-fold in triple negative vs. luminal cells. This suggested that the differential detection of these glycosites in triple negative CM samples may have been attributable to differences in relative protein abundances. In contrast, more than half of the triple negative-specific candidates could not have been predicted from the mRNA expression data, as there was no difference in mRNA abundances between the luminal and triple negative subsets. The identification of these glycosites may have been driven by alterations in the protein glycosylation machinery of triple negative cell lines. To address this possibility, we looked for differences in mRNA levels of the transferases that add fucose (recognized by AAL), and sialic acid (recognized by SNA). The results are shown in Supplementary Table 5. Two fucosyltransferases and 8 sialyltransferases were differentially expressed, either up or downregulated, in triple negative vs. luminal cell lines. Given that we observed both gains and losses of enzymatic activity, it is difficult to predict, in structural terms, the net consequences of these changes. However, our glycosite data are empirical evidence of subtype-specific glycosylation patterns in breast cancer.

Disease relevance of biomarker scaffolds

Initial inspection of the 100 triple negative-specific candidates showed that many targets were derived from glycoproteins that are involved in cancer-relevant processes. To more fully explore this correlation, we performed pathway analyses using two bioinformatics resources: Kyoto Encyclopedia of Genes and Genomes (KEGG) and Ingenuity (IPA). However, the programs recognized only small portions of the dataset, together matching 38% of the total proteins (Supplementary Tables 6 and 7), and most of the results were driven by only a few molecules, *e.g.*, integrins. As an alternative, literature searches enabled assignment of biological functions to 90% of the putative triple negative-specific glycoproteins. Three prominent, interrelated themes emerged—38% of the targets were up- or downstream components of the TGF β pathway; 21% were involved in ECM remodeling; and at least 18% were proteinases or proteolytic targets. Minor recurring associations included the epithelial to mesenchymal transition (EMT, 9%) and bone morphogenic protein signaling (6%).

TGF β signaling governs important aspects of ECM remodeling and proteinase activities. Through the synthesis, cross-linking, and degradation of a variety of protein and carbohydrate matrix components, the composition and tensile strength of the ECM are modulated, both of which dramatically influence the behavior of surrounding cells^{42, 43}. With respect to cancer, these activities are strongly associated with increased migration and invasion. TGF β is also considered to be a central mediator of EMT, through both canonical (*i.e.*, Smad-dependent) and non-canonical (*e.g.*, PI3K and MAPK) pathways⁴⁴. Cells undergoing EMT lose apical-basal polarity and stabilizing adhesive epithelial interactions in exchange for the acquisition of a more migratory mesenchymal phenotype. These changes can lead to cell invasion and metastasis, functions that have been linked to TGF β

activity^{45, 46}. Thus, as a group, the putative triple negative-specific targets we identified were derived from proteins with striking functional similarities and disease relevance⁴⁷. It is possible that these biomarker candidates may also suggest subtype-specific clinical targets, which currently do not exist for triple negative breast cancer^{18, 19}.

Clinical relevance of putative biomarker targets

The heterogeneous nature of breast cancer is widely accepted¹³. Tumor subtyping is commonly based on immunohistochemical analyses of tissue sections cut from biopsies to profile expression of a marker panel—ER, PR, HER2, cytokeratin 5/6 and epidermal growth factor receptor. Increasingly, clinicians are using this information to determine prognoses and optimize treatment⁴⁸. For example, the risk prediction tool Adjuvant!Online (www.adjuvantonline.com) can be used to identify the patients who will benefit most from postoperative treatment(s). Although immunohistology-based diagnoses are changing the clinical oncology landscape and improving patient outcomes, there remains much room for advancement. Currently, subtype diagnoses require identification of a lesion, and an invasive procedure to obtain a biopsy. Therefore, the need for circulating biomarkers that serve as sentinels of breast cancer and enable subtyping remains great.

In this context, our biomarker discovery method used cancer cell line CM, *i.e.*, the secretome, as the starting material to identify candidate glycoproteins that carried putative subtype-specific *N*-glycosites. For the enrichment step, we used lectin capture at the glycopeptide, rather than glycoprotein level. This approach gives more information, in terms of glycan composition and location along the peptide backbone, than other commonly used related methods (*e.g.*, lectin chromatography at the glycoprotein level, and hydrazide- or boronic acid-mediated chemical capture of glycoproteins/glycopeptides)²⁴. Accordingly, we interrogated a largely unexplored biomarker discovery space. This theory is substantiated by the fact that only four of the targets that we identified were among the 150 most abundant plasma proteins as described by Hortin *et al.*⁴⁹. Furthermore, only 52 of the targets were among the recently published high-confidence human plasma proteome that included estimated protein concentrations⁵⁰. Of those found in this dataset, 73% were predicted to be <50 ng/mL, while 40% were likely to be <10 ng/mL, reasonably low background levels against which to observe circulating disease-derived signals. As additional support for this concept, only six of the putative triple negative-specific *N*-glycosites from five glycoproteins were found in a previous study in which we used the same workflows and AAL or SNA chromatography to analyze a sample of NIST pooled human plasma from 100 healthy individuals²⁴. These included glycosites from CD109, CD44, clusterin, extracellular matrix protein 1, and pigment epithelium-derived factor.

In summary, the workflow that we developed could serve as a blueprint for biomarker discovery. In this paradigm, an initial candidate list is developed using an easily obtained renewable material, such as cell line CM, rather than valuable, and often difficult to obtain, clinical samples such as plasma or serum. As studies that employ targeted enrichment strategies are considerably more sensitive than shotgun proteomics methods, the ability to generate a candidate biomarker list from a biologically-relevant source significantly improves the chances of success during the subsequent verification stage⁵¹. This method may be especially useful for diseases, such as ovarian cancer, for which the cell type of origin is uncertain and, consequently, it is difficult to choose control samples^{52, 53}. A limitation of the method is that O-linked and intact N-linked glycopeptides are not analyzed due to the absence of universal enzymes to remove carbohydrates and the lack of sufficiently powerful software for rapid identifications, respectively. However, we do not view this as a liability. This workflow was designed as a high-throughput platform to generate biomarker candidates for subsequent verification by MRM. In general, due to heterogeneity, endogenous glycopeptides make poor MRM targets. By contrast, our method

yielded a list of putative biomarker targets for direct follow up in clinical samples, and is easily accessible to any laboratory performing proteomics. Indeed, several groups have recently employed similar methods to identify candidate biomarkers of various cancers including prostate, colon, thyroid and breast^{54–57}. Interestingly, a few of the biomarkers that we identified were also observed in the latter study, suggesting that this general approach is reproducible and robust⁵⁴. Finally, this workflow is well suited to the development of a multiplexed clinical assay, analogous to a reverse protein array approach, with antibody capture as the first step and lectin binding as the second.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ms. Tiffany Sham for excellent assistance formatting tables. This work was supported by an NCRR shared instrumentation grant S10 RR024615 (BWG) and by grants from the National Cancer Institute, U24 CA126477 (SJF) and a U24 Subcontract (BWG) that are part of the NCI Clinical Proteomic Technologies for Cancer initiative (<http://proteomics.cancer.gov>). Additional support was provided by the Director, Office of Science, Office of Biological & Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, by the National Institutes of Health, National Cancer Institute grants P50 CA 58207, the U54 CA 112970, the U24 CA 126477 and the NIH NHGRI U24 CA 126551 for JWG. A portion of the mass spectrometric analyses was performed in the UCSF Sandler-Moore Mass Spectrometry Core Facility, which acknowledges support from the Sandler Family Foundation, the Gordon and Betty Moore Foundation, and NIH/NCI Cancer Center Support Grant P30 CA082103. OLG is supported by the Canadian Institutes of Health Research and the Stand Up To Cancer-American Association for Cancer Research Dream Team Translational Cancer Research Grant SU2C-AACR-DT0409.

References

1. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, Regnier FE, Gibson BW, Fisher SJ. Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clin Chem*. 2010; 56:223–236. [PubMed: 19959616]
2. Hart GW, Copeland RJ. Glycomics hits the big time. *Cell*. 2010; 143(5):672–6. [PubMed: 21111227]
3. Clowers BH, Dodds ED, Seipert RR, Lebrilla CB. Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and Fourier transform ion cyclotron resonance mass spectrometry. *J Proteome Res*. 2007; 6(10):4032–40. [PubMed: 17824634]
4. Duffy MJ, Evoy D, McDermott EW. CA 15–3: uses and limitation as a biomarker for breast cancer. *Clin Chim Acta*. 2010; 411(23–24):1869–74. [PubMed: 20816948]
5. Orntoft TF, Vestergaard EM. Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis*. 1999; 20(2):362–71. [PubMed: 10197444]
6. Hammarstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol*. 1999; 9(2):67–81. [PubMed: 10202129]
7. Meany DL, Zhang Z, Sokoll LJ, Zhang H, Chan DW. Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. *J Proteome Res*. 2009; 8(2):613–9. [PubMed: 19035787]
8. Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. *J Clin Pathol*. 2005; 58(3):308–12. [PubMed: 15735166]
9. Witz IP. The selectin-selectin ligand axis in tumor progression. *Cancer Metastasis Rev*. 2008; 27(1):19–30. [PubMed: 18180878]
10. Rosen SD. Ligands for L-selectin: homing, inflammation, and beyond. *Annu Rev Immunol*. 2004; 22:129–56. [PubMed: 15032576]
11. Perou CM, Borresen-Dale AL. Systems Biology and Genomics of Breast Cancer. *Cold Spring Harb Perspect Biol*. 2010

12. O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, Dressler LG, Geradts J, Millikan RC. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res.* 2010; 16(24):6100–10. [PubMed: 21169259]
13. Espinosa E, Vara JA, Navarro IS, Gamez-Pozo A, Pinto A, Zamora P, Redondo A, Feliu J. Gene profiling in breast cancer: Time to move forward. *Cancer Treat Rev.* 2011
14. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 2010; 12(5):R68. [PubMed: 20813035]
15. Toft DJ, Cryns VL. Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. *Mol Endocrinol.* 2011; 25(2):199–211. [PubMed: 20861225]
16. Abramson V, Arteaga CL. New strategies in HER2-overexpressing breast cancer: Many combinations of targeted drugs available. *Clin Cancer Res.* 2011
17. McDermott U, Settleman J. Personalized cancer therapy with selective kinase inhibitors: an emerging paradigm in medical oncology. *J Clin Oncol.* 2009; 27(33):5650–9. [PubMed: 19858389]
18. Yagata H, Kajiura Y, Yamauchi H. Current strategy for triple-negative breast cancer: appropriate combination of surgery, radiation, and chemotherapy. *Breast Cancer.* 2011
19. Pal SK, Childs BH, Pegram M. Triple negative breast cancer: unmet medical needs. *Breast Cancer Res Treat.* 2011; 125(3):627–36. [PubMed: 21161370]
20. Lapuk A, Marr H, Jakkula L, Pedro H, Bhattacharya S, Purdom E, Hu Z, Simpson K, Pachter L, Durinck S, Wang N, Parvin B, Fontenay G, Speed T, Garbe J, Stampfer M, Bayandorian H, Dorton S, Clark TA, Schweitzer A, Wyrobek A, Feiler H, Spellman P, Conboy J, Gray JW. Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res.* 2010; 8(7):961–74. [PubMed: 20605923]
21. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell.* 2006; 10(6):515–27. [PubMed: 17157791]
22. Korkola J, Gray JW. Breast cancer genomes--form and function. *Curr Opin Genet Dev.* 2010; 20(1):4–14. [PubMed: 20060285]
23. Kuo WL, Das D, Ziyad S, Bhattacharya S, Gibb WJ, Heiser LM, Sadanandam A, Fontenay GV, Hu Z, Wang NJ, Bayani N, Feiler HS, Neve RM, Wyrobek AJ, Spellman PT, Marton LJ, Gray JW. A systems analysis of the chemosensitivity of breast cancer cells to the polyamine analogue PG-11047. *BMC Med.* 2009; 7:77. [PubMed: 20003408]
24. Drake PM, Schilling B, Niles RK, Braten M, Johansen E, Liu H, Lerch M, Sorensen DJ, Li B, Allen S, Hall SC, Witkowska HE, Regnier FE, Gibson BW, Fisher SJ. A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Anal Biochem.* 2011; 408(1):71–85. [PubMed: 20705048]
25. Yingling JM, Blanchard KL, Sawyer JS. Development of TGF-beta signalling inhibitors for cancer therapy. *Nat Rev Drug Discov.* 2004; 3(12):1011–22. [PubMed: 15573100]
26. Janatpour MJ, McMaster MT, Genbacev O, Zhou Y, Dong J, Cross JC, Israel MA, Fisher SJ. Id-2 regulates critical aspects of human cytotrophoblast differentiation, invasion and migration. *Development.* 2000; 127(3):549–58. [PubMed: 10631176]
27. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics.* 2007; 6(12):2212–29. [PubMed: 17939991]
28. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics.* 2007; 6(9):1638–55. [PubMed: 17533153]
29. Krokhn OV, Antonovici M, Ens W, Wilkins JA, Standing KG. Deamidation of - Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem.* 2006; 78(18):6645–50. [PubMed: 16970346]

30. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol.* 1999; 17(7):676–82. [PubMed: 10404161]
31. Edgington, ES. Randomization tests. 3. Marcel-Dekker; New York: 1995.
32. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010; 26(7):966–8. [PubMed: 20147306]
33. Griffith M, Griffith OL, Mwenifumbo J, Goya R, Morrissy AS, Morin RD, Corbett R, Tang MJ, Hou YC, Pugh TJ, Robertson G, Chittaranjan S, Ally A, Asano JK, Chan SY, Li HI, McDonald H, Teague K, Zhao Y, Zeng T, Delaney A, Hirst M, Morin GB, Jones SJ, Tai IT, Marra MA. Alternative expression analysis by RNA sequencing. *Nat Methods.* 2010; 7(10):843–7. [PubMed: 20835245]
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological).* 1995; 57(1):289–300.
35. Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, Spiegelman CH, Zimmerman LJ, Ham AJ, Keshishian H, Hall SC, Allen S, Blackman RK, Borchers CH, Buck C, Cardasis HL, Cusack MP, Dodder NG, Gibson BW, Held JM, Hiltke T, Jackson A, Johansen EB, Kinsinger CR, Li J, Mesri M, Neubert TA, Niles RK, Pulsipher TC, Ransohoff D, Rodriguez H, Rudnick PA, Smith D, Tabb DL, Tegeler TJ, Variyath AM, Vega-Montoto LJ, Wahlander A, Waldemarson S, Wang M, Whiteaker JR, Zhao L, Anderson NL, Fisher SJ, Liebler DC, Paulovich AG, Regnier FE, Tempst P, Carr SA. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol.* 2009; 27(7):633–41. [PubMed: 19561596]
36. Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM, Kilpatrick LE, Billheimer DD, Blackman RK, Cardasis HL, Carr SA, Clauser KR, Jaffe JD, Kowalski KA, Neubert TA, Regnier FE, Schilling B, Tegeler TJ, Wang M, Wang P, Whiteaker JR, Zimmerman LJ, Fisher SJ, Gibson BW, Kinsinger CR, Mesri M, Rodriguez H, Stein SE, Tempst P, Paulovich AG, Liebler DC, Spiegelman C. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res.* 2010; 9(2):761–76. [PubMed: 19921851]
37. Wu ZS, Wu Q, Yang JH, Wang HQ, Ding XD, Yang F, Xu XC. Prognostic significance of MMP-9 and TIMP-1 serum and tissue expression in breast cancer. *Int J Cancer.* 2008; 122(9):2050–6. [PubMed: 18172859]
38. Wang Y, Ao X, Vuong H, Konanur M, Miller FR, Goodison S, Lubman DM. Membrane glycoproteins associated with breast tumor cell progression identified by a lectin affinity approach. *J Proteome Res.* 2008; 7(10):4313–25. [PubMed: 18729497]
39. Baricevic I, Masnikosa R, Lagundzin D, Golubovic V, Nedic O. Alterations of insulin-like growth factor binding protein 3 (IGFBP-3) glycosylation in patients with breast tumours. *Clin Biochem.* 2010; 43(9):725–31. [PubMed: 20307522]
40. Bast RC Jr, Xu FJ, Yu YH, Barnhill S, Zhang Z, Mills GB. CA 125: the past and the future. *Int J Biol Markers.* 1998; 13(4):179–87. [PubMed: 10228898]
41. Yin BW, Lloyd KO. Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16. *J Biol Chem.* 2001; 276(29):27371–5. [PubMed: 11369781]
42. Yu H, Mouw JK, Weaver VM. Forcing form and function: biomechanical regulation of tumor evolution. *Trends Cell Biol.* 2011; 21(1):47–56. [PubMed: 20870407]
43. Rowe RG, Weiss SJ. Navigating ECM barriers at the invasive front: the cancer cell-stroma interface. *Annu Rev Cell Dev Biol.* 2009; 25:567–95. [PubMed: 19575644]
44. Vitoria-Petit AM, Wrana JL. The TGFbeta-Par6 polarity pathway: linking the Par complex to EMT and breast cancer progression. *Cell Cycle.* 2010; 9(4):623–4. [PubMed: 20107326]
45. Barcellos-Hoff MH, Akhurst RJ. Transforming growth factor-beta in breast cancer: too much, too late. *Breast Cancer Res.* 2009; 11(1):202. [PubMed: 19291273]
46. Bergers G, Javaherian K, Lo KM, Folkman J, Hanahan D. Effects of angiogenesis inhibitors on multistage carcinogenesis in mice. *Science.* 1999; 284(5415):808–12. [PubMed: 10221914]

47. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144(5):646–74. [PubMed: 21376230]
48. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C, Heikkila P, Heikkinen T, Nevanlinna H, Akslen LA, Begin LR, Foulkes WD, Couch FJ, Wang X, Cafourek V, Olson JE, Baglietto L, Giles GG, Severi G, McLean CA, Southey MC, Rakha E, Green AR, Ellis IO, Sherman ME, Lissowska J, Anderson WF, Cox A, Cross SS, Reed MW, Provenzano E, Dawson SJ, Dunning AM, Humphreys M, Easton DF, Garcia-Closas M, Caldas C, Pharoah PD, Huntsman D. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med*. 2010; 7(5):e1000279. [PubMed: 20520800]
49. Hortin GL, Sviridov D, Anderson NL. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin Chem*. 2008; 54(10):1608–16. [PubMed: 18687737]
50. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmstrom J, Ossola R, Watts JD, Lin B, Zhang H, Moritz RL, Aebersold RH. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics*. 2011
51. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*. 2006; 24(8):971–83. [PubMed: 16900146]
52. Lengyel E. Ovarian cancer development and metastasis. *Am J Pathol*. 2010; 177(3):1053–64. [PubMed: 20651229]
53. Vang R, Shih Ie M, Kurman RJ. Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems. *Adv Anat Pathol*. 2009; 16(5):267–82. [PubMed: 19700937]
54. Ahn Y, Kang UB, Kim J, Lee C. Mining of serum glycoproteins by an indirect approach using cell line secretome. *Mol Cells*. 2010; 29(2):123–30. [PubMed: 20016938]
55. Arcinas A, Yen TY, Kebebew E, Macher BA. Cell surface and secreted protein profiles of human thyroid cancer cell lines reveal distinct glycoprotein patterns. *J Proteome Res*. 2009; 8(8):3958–68. [PubMed: 19530676]
56. Rangiah K, Tippornwong M, Sangar V, Austin D, Tetreault MP, Rustgi AK, Blair IA, Yu KH. Differential secreted proteome approach in murine model for candidate biomarker discovery in colon cancer. *J Proteome Res*. 2009; 8(11):5153–64. [PubMed: 19769411]
57. Sardana G, Jung K, Stephan C, Diamandis EP. Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. *J Proteome Res*. 2008; 7(8):3329–38. [PubMed: 18578523]
58. Akasaka-Manyá K, Manyá H, Sakurai Y, Wojczyk BS, Spitalnik SL, Endo T. Increased bisecting and core-fucosylated N-glycans on mutant human amyloid precursor proteins. *Glycoconj J*. 2008; 25(8):775–86. [PubMed: 18521746]
59. Nakagawa K, Kitazume S, Oka R, Maruyama K, Saido TC, Sato Y, Endo T, Hashimoto Y. Sialylation enhances the secretion of neurotoxic amyloid-beta peptides. *J Neurochem*. 2006; 96(4): 924–33. [PubMed: 16412100]
60. Garrigue-Antar L, Hartigan N, Kadler KE. Post-translational modification of bone morphogenetic protein-1 is required for secretion and stability of the protein. *J Biol Chem*. 2002; 277(45):43327–34. [PubMed: 12218058]
61. Wolf B. Biotinidase Deficiency: New Directions and Practical Concerns. *Curr Treat Options Neurol*. 2003; 5(4):321–328. [PubMed: 12791199]
62. Ciolczyk-Wierzbicka D, Amoresano A, Casbarra A, Hoja-Lukowicz D, Litynska A, Laidler P. The structure of the oligosaccharides of N-cadherin from human melanoma cell lines. *Glycoconj J*. 2004; 20(7–8):483–92. [PubMed: 15316281]
63. Takahashi T, Schmidt PG, Tang J. Novel carbohydrate structures of cathepsin B from porcine spleen. *J Biol Chem*. 1984; 259(10):6059–62. [PubMed: 6725247]
64. Dimitroff CJ, Lee JY, Rafii S, Fuhlbrigge RC, Sackstein R. CD44 is a major E-selectin ligand on human hematopoietic progenitor cells. *J Cell Biol*. 2001; 153(6):1277–86. [PubMed: 11402070]

65. Elliott MM, Kardana A, Lustbader JW, Cole LA. Carbohydrate and peptide structure of the alpha- and beta-subunits of human chorionic gonadotropin from normal and aberrant pregnancy and choriocarcinoma. *Endocrine*. 1997; 7(1):15–32. [PubMed: 9449027]
66. Kapron JT, Hilliard GM, Lakins JN, Tenniswood MP, West KA, Carr SA, Crabb JW. Identification and characterization of glycosylation sites in human serum clusterin. *Protein Sci*. 1997; 6(10):2120–33. [PubMed: 9336835]
67. Goldberg M, Peshkovsky C, Shifteh A, Al-Awqati Q. mu-Protocadherin, a novel developmentally regulated protocadherin with mucin-like domains. *J Biol Chem*. 2000; 275(32):24622–9. [PubMed: 10801787]
68. Hirnle L, Katnik-Prastowska I. Amniotic fibronectin fragmentation and expression of its domains, sialyl and fucosyl glycotopes associated with pregnancy complicated by intrauterine infection. *Clin Chem Lab Med*. 2007; 45(2):208–14. [PubMed: 17311510]
69. Miyamae T, Marinov AD, Sowders D, Wilson DC, Devlin J, Boudreau R, Robbins P, Hirsch R. Follistatin-like protein-1 is a novel proinflammatory molecule. *J Immunol*. 2006; 177(7):4758–62. [PubMed: 16982916]
70. Hyuga M, Itoh S, Kawasaki N, Ohta M, Ishii A, Hyuga S, Hayakawa T. Analysis of site-specific glycosylation in recombinant human follistatin expressed in Chinese hamster ovary cells. *Biologicals*. 2004; 32(2):70–7. [PubMed: 15454184]
71. Ohgomori T, Funatsu O, Nakaya S, Morita A, Ikekita M. Structural study of the N-glycans of intercellular adhesion molecule-5 (telencephalin). *Biochim Biophys Acta*. 2009; 1790(12):1611–23. [PubMed: 19733219]
72. Pochee E, Litynska A, Amoresano A, Casbarra A. Glycosylation profile of integrin alpha 3 beta 1 changes with melanoma progression. *Biochim Biophys Acta*. 2003; 1643(1–3):113–23. [PubMed: 14654234]
73. Newburg DS, Peterson JA, Ruiz-Palacios GM, Matson DO, Morrow AL, Shults J, Guerrero ML, Chaturvedi P, Newburg SO, Scallan CD, Taylor MR, Ceriani RL, Pickering LK. Role of human-milk lactadherin in protection against symptomatic rotavirus infection. *Lancet*. 1998; 351(9110):1160–4. [PubMed: 9643686]
74. Stimson E, Hope J, Chong A, Burlingame AL. Site-specific characterization of the N-linked glycans of murine prion protein by high-performance liquid chromatography/electrospray mass spectrometry and exoglycosidase digestions. *Biochemistry*. 1999; 38(15):4885–95. [PubMed: 10200178]
75. Sato S, Rahemtulla F, Prince CW, Tomana M, Butler WT. Acidic glycoproteins from bovine compact bone. *Connect Tissue Res*. 1985; 14(1):51–64. [PubMed: 2934214]
76. Nakahara Y, Miyata T, Hamuro T, Funatsu A, Miyagi M, Tsunasawa S, Kato H. Amino acid sequence and carbohydrate structure of a recombinant human tissue factor pathway inhibitor expressed in Chinese hamster ovary cells: one N- and two O-linked carbohydrate chains are located between Kunitz domains 2 and 3 and one N-linked carbohydrate chain is in Kunitz domain 2. *Biochemistry*. 1996; 35(20):6450–9. [PubMed: 8639592]
77. Brunner AM, Gentry LE, Cooper JA, Purchio AF. Recombinant type 1 transforming growth factor beta precursor produced in Chinese hamster ovary cells is glycosylated and phosphorylated. *Mol Cell Biol*. 1988; 8(5):2229–32. [PubMed: 3164441]
78. Ploug M, Rahbek-Nielsen H, Nielsen PF, Roepstorff P, Dano K. Glycosylation profile of a recombinant urokinase-type plasminogen activator receptor expressed in Chinese hamster ovary cells. *J Biol Chem*. 1998; 273(22):13933–43. [PubMed: 9593742]

Synopsis

We used a lectin chromatography/MS-based approach to screen conditioned medium from a panel of luminal (less aggressive) and triple negative (more aggressive) breast cancer cell lines. The samples were fractionated using lectins that recognize fucose and sialic acid. In total, we identified 1011 glycosites from 533 glycoproteins. Statistical analyses suggested that a number of these glycosites were triple negative-specific and thus potential biomarkers for this tumor subtype.

Fig. 1. Breast cancer cell lines have a complex repertoire of SNA-reactive glycoproteins and exhibit cell surface staining with this lectin

(A) Lysates from a panel of 8 breast cancer cell lines, which included triple negative (1–6) and luminal (7, 8) subtypes, were electrophoretically separated, transferred to nitrocellulose, and probed with SNA. Lane 1. MDAMB468, 2. HCC38, 3. HCC1500, 4. HS578T, 5. MDAMB157, 6. MDAMB231, 7. T47D, 8. UCC812. (B) Non-permeabilized HS578T cells were stained with fluorescein-conjugated SNA and imaged by fluorescence microscopy (magnification 60x).

Fig. 2. The experimental workflow

(A) CM samples from breast cancer cell lines established from five luminal and 5 triple negative tumors were prepared in one laboratory, then distributed to 3 sites. (B) Each group separated the 10 CM samples, in duplicate, by AAL or SNA chromatography, which generated 40 fractions. The samples were deglycosylated using PNGaseF and analyzed in duplicate by LC-MS/MS, yielding a total of 80 MS/MS data sets per site. (C) Files were transferred to a central location for bioinformatic analyses.

Fig. 3. Diagrammatic summary of the glycosite (glycoprotein) enrichment data according to lectin type (AAL vs. SNA) and CM samples (luminal vs. triple negative) showed distinct and overlapping specificities

(A) The intersecting circles depict the total number of *N*-glycosites (glycoproteins) captured by each lectin. (B and C) Venn diagrams illustrating the chromatographic separation of luminal (LUM) and triple negative (TN) CM samples.

Fig. 4. Lectin capture resulted in significant glycopeptide enrichment

The percent enrichment for the separations performed using AAL (*left*) or SNA (*right*) at Sites M (*top*), X (*middle*), and S (*bottom*). The dark line indicates the median; the box depicts the first and third quartiles; the whiskers show the minimum and maximum values observed. Sites M and X acquired data using QSTAR Elite instruments, while Site S used an Orbitrap mass spectrometer.

Fig. 5. Nearly 90% of identified glycoproteins resided in the plasma membrane or extracellular compartments

A portion (241/560) of the identified glycoproteins were annotated in the cellular component of Gene Ontology. Of these, the great majority were cell surface or secreted molecules.

Triple negative-specific glycosites (glycoproteins)

Fig. 6. Putative triple negative-specific glycosites (glycoproteins) enriched by AAL or SNA
The criteria applied were detection in 3 triple negative and 0 luminal cell line CMs.

Table 1

Luminal and triple-negative breast cancer cell lines.

Cell line ^a	Tumor subtype	ER ^b	PR ^c	HER2 ^d	Diagnosis ^e
SKBR3	Luminal	–	–	Yes	Adenocarcinoma
SUM52 PE	Luminal	+	–	No	Carcinoma
MDAMB175	Luminal	+	–	No	IDC
UACC 812	Luminal	+	–	Yes	IDC
MDAMA361	Luminal	+	–	Yes	Adenocarcinoma
MDA468	Basal A	–	–	No	Adenocarcinoma
BT549	Claudin-low	–	–	No	IDC, papillary
HS5787	Claudin-low	–	–	No	IDC
MDAMB 231	Claudin-low	–	–	No	Adenocarcinoma
HCC38	Claudin-low	–	–	No	Ductal carcinoma

^aThis table is populated with information from Neve et al., 2006.

^{b,c} Estrogen (ER) or progesterone receptor (PR) expression.

^d Human epidermal growth factor receptor 2 (HER2) overexpression.

^e Invasive ductal carcinoma (IDC).

Table 2

Putative triple negative-specific glycosites captured by AAL and SNA.

Gene name	Entry name	Accession Number ^d	Glycoprotein	Glycosite(s)	Observed TN AAL ^b	Observed TN SNA ^b	Function ^c	Known Fucosylation or Sialylation ^d
NT5E	5NTD_HUMAN	P21589	5'-nucleotidase	N333	3	2	Nucleotidase	
APP	A4_HUMAN	P05067	Amyloid beta A4 protein	N571	3	1	Adhesion molecule	Yes (58, 59)
ANPEP	AMPN_HUMAN	P15144	Aminopeptidase N	N265	0	3	Protease	
ANTXR1	ANTR1_HUMAN	Q96EC6	Anthrax toxin receptor 1	N184	4	2	Adhesion molecule	
B3GNT2	B3GN2_HUMAN	B3GNT2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	N173	3	1	Glycosyltransferase	
BMP1	BMP1_HUMAN	P13497	Bone morphogenetic protein 1	N91	4	1	Protease	Yes (60)
BMP1	BMP1_HUMAN	P13497	Bone morphogenetic protein 1	N142	3	1	Protease	Yes (60)
BTD	BTD_HUMAN	P43251	Biotinidase	N150	4	3	Metabolism	Yes (61)
CDH2	CADH2_HUMAN	P19022	Cadherin-2	N325	3	0	Adhesion molecule	Yes (62)
CDH2	CADH2_HUMAN	P19022	Cadherin-2	N402	3	0	Adhesion molecule	Yes (62)
CDH2	CADH2_HUMAN	P19022	Cadherin-2	N692	3	3	Adhesion molecule	Yes (62)
CTSB	CATB_HUMAN	P07858	Cathepsin B	N38	5	4	Protease	Yes (63)
CTSL1	CATL1_HUMAN	P07711	Cathepsin L1	N221	4	3	Protease	
CTSL2	CATL2_HUMAN	O60911	Cathepsin L2	N221	3	2	Protease	
CCDC80	CCD80_HUMAN	Q8R2G6	Coiled-coil domain-containing protein 80	N667	3	1	Adhesion molecule	
CCDC80	CCD80_HUMAN	Q8R2G6	Coiled-coil domain-containing protein 80	N668	4	3	Adhesion molecule	
CD109	CD109_HUMAN	Q6YHK3	CD109 antigen ^e	N68	3	1	TGF-beta pathway	
CD109	CD109_HUMAN	Q6YHK3	CD109 antigen ^e	N397	4	0	TGF-beta pathway	
CD44	CD44_HUMAN	P16070	CD44 antigen ^e	N25	3	5	Adhesion molecule	Yes (64)
CGB1	CGB1_HUMAN	A6NKQ9	Choriogonadotropin subunit beta variant 1	N63	3	0	Hormone	Yes (65)
CLIC1	CLIC1_HUMAN	O00299	Chloride intracellular channel protein 1	N42	4	2	Ion channel	
CLU	CLUS_HUMAN	P10909	Clusterin ^e	N354	4	3	Receptor	Yes (66)
COL1A1	CO1A1_HUMAN	P02452	Collagen alpha-1 (I) chain	N1365	2	3	ECM	
COL5A1	CO5A1_HUMAN	P20908	Collagen alpha-1 (V) chain	N176	3	4	ECM	
COL6A1	CO6A1_HUMAN	P12109	Collagen alpha-1 (VI) chain	N804	3	3	ECM	

Gene name	Entry name	Accession Number ^a	Glycoprotein	Glycosite(s)	Observed TN AAL ^b	Observed TN SNA ^b	Function ^c	Known Fucosylation or Sialylation ^d
COL6A2	CO6A2_HUMAN	P12110	Collagen alpha-2 (V1) chain	N140	2	3	ECM	
COL6A2	CO6A2_HUMAN	P12110	Collagen alpha-2 (V1) chain	N785	3	3	ECM	
COL12A1	COCA1_HUMAN	Q99715	Collagen alpha-1 (XII) chain	N2679	4	4	ECM	
COL18A1	COIA1_HUMAN	P39060	Collagen alpha-1 (XVIII) chain	N926	3	0	ECM	
CPVL	CPVL_HUMAN	Q9H3G5	Probable serine carboxypeptidase CPVL	N346	3	1	Protease	
CRIM1	CRIM1_HUMAN	Q9NZV1	Cysteine-rich motor neuron 1 protein	N71	3	3	Receptor	
CRTAP	CRTAP_HUMAN	O75718	Cartilage-associated protein	N87	3	2	ECM	
DCBLD1	DCBD1_HUMAN	Q8N8Z6	Discoidin, CUB and LCCL domain-containing protein 1	N124	4	2	Unknown	
DKK3	DKK3_HUMAN	Q9UBP4	Dickkopf-related protein 3	N96	3	1	Wnt signaling pathway	
ECE1	ECE1_HUMAN	P42892	Endothelin-converting enzyme 1	N166	3	0	Protease	
ECM1	ECM1_HUMAN	Q16610	Extracellular matrix protein 1 ^e	N444	3	3	Angiogenesis	
EXT1	EXT1_HUMAN	Q16394	Exostosin-1	N330	4	2	GAG synthesis	
EXT2	EXT2_HUMAN	Q93063	Exostosin-2	N288	3	0	GAG synthesis	
FAT1	FAT1_HUMAN	Q14517	Protocadherin Fat 1	N2328	3	0	Adhesion molecule	Yes (67)
FBN1	FBN1_HUMAN	P35555	Fibrillin-1	N1581	4	3	TGF-beta pathway	
FBN1	FBN1_HUMAN	P35555	Fibrillin-1	N448 and N2767	4	4	TGF-beta pathway	
FN1	FINC_HUMAN	P02751	Fibronectin	N430	3	3	ECM	Yes (68)
FSTL1	FSTL1_HUMAN	Q12841	Follistatin-related protein 1	N175	4	5	Cell growth Differentiation	Yes (69)
FSTL1	FSTL1_HUMAN	Q12841	Follistatin-related protein 1	N180	3	5	Cell growth, Differentiation	Yes (69)
FSTL3	FSTL3_HUMAN	O95633	Follistatin-related protein 3	N215	4	3	TGF-beta pathway	
FST	FST_HUMAN	P19883	Follistatin	N288	3	2	Hormonal regulation	Yes (70)
SERPINE2	GDN_HUMAN	P07093	Glia-derived nexin	N118	4	3	Protease inhibitor	
GCNT2	GNT2A_HUMAN	Q8N0V5	N-acetylglucosaminide beta-1,6-N-acetylglucosaminyl-transferase, isoform A	N41	3	2	Glycosyltransferase	
GPC1	GPC1_HUMAN	P35052	Glypican-1	N79	5	4	GAG	
GRN	GRN_HUMAN	P28799	Granulins	N236	5	4	Cytokine	
HSPA13	HSP13_HUMAN	P48723	Heat shock 70 kDa protein 13	N184	4	3	ATPase	
IGFBP3	IBP3_HUMAN	P17936	Insulin-like growth factor-binding protein 3	N199	5	4	Cell growth, Differentiation	
ICAM5	ICAM5_HUMAN	Q9UMF0	Intercellular adhesion molecule 5	N646	3	2	Adhesion molecule	Yes (71)

Gene name	Entry name	Accession Number ^a	Glycoprotein	Glycosite(s)	Observed TN AAL ^b	Observed TN SNA ^b	Function ^c	Known Fucosylation or Sialylation ^d
ITGA3	ITA3_HUMAN	P26006	Integrin alpha-3	N265	3	0	Adhesion molecule	Yes (72)
ITGA5	ITA5_HUMAN	P08648	Integrin alpha-5	N868	3	1	Adhesion molecule	
ITGB1	ITB1_HUMAN	P05556	Integrin beta-1	N520	4	3	Adhesion molecule	Yes (72)
ITGB1	ITB1_HUMAN	P05556	Integrin beta-1	N669	4	3	Adhesion molecule	Yes (72)
JAG1	JAG1_HUMAN	P78504	Protein jagged-1	N217	4	0	Cell growth, Differentiation	
LAMC1	LAMC1_HUMAN	P11047	Laminin subunit gamma-1	N1205	4	3	ECM	
LAMC1	LAMC1_HUMAN	P11047	Laminin subunit gamma-1	N1395	5	3	ECM	
LIF	LIF_HUMAN	P09056	Leukemia inhibitory factor	N85	3	0	Cell growth, Differentiation	
LOXL2	LOXL2_HUMAN	Q9Y4K0	Lysyl oxidase homolog 2	N288	4	3	ECM cross-linking	
LOX	LYOX_HUMAN	P28300	Protein-lysine 6-oxidase	N81	4	2	ECM cross-linking	
LOX	LYOX_HUMAN	P28300	Protein-lysine 6-oxidase	N144	4	1	ECM cross-linking	
MET	MET_HUMAN	P08581	Hepatocyte growth factor receptor	N106	3	0	Cell growth, Differentiation	
MFGE8	MFGM_HUMAN	Q08431	Lactadherin	N325	3	4	Tissue homeostasis	Yes (73)
MICA	MICA_HUMAN	Q29983	MHC class I polypeptide-related sequence A	N79	4	3	Immune regulator	
MRC2	MRC2_HUMAN	Q9UBG0	C-type mannose receptor 2	N497	4	1	ECM remodeling	
OLFML3	OLF3_HUMAN	Q9NRN5	Olfactomedin-like protein 3	N177	4	3	Development	
LEPRE1	P3H1_HUMAN	Q32P28	Prolyl 3-hydroxylase 1	N540	4	4	GAG	
SERPINF1	PEDE_HUMAN	P36955	Pigment epithelium-derived factor ^e	N285	1	4	Cell growth, Differentiation	
PLOD2	PLOD2_HUMAN	O00469	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	N63	4	1	ECM cross-linking	
PLOD3	PLOD3_HUMAN	O60568	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	N548	3	0	ECM cross-linking	
PLTP	PLTP_HUMAN	P55058	Phospholipid transfer protein	N143	3	3	Lipid metabolism	
PLTP	PLTP_HUMAN	P55058	Phospholipid transfer protein	N398	3	3	Lipid metabolism	
POSTN	POSTN_HUMAN	Q15063	Periostin	N599	4	3	Adhesion molecule	
PPGB	PPGB_HUMAN	P10619	Lysosomal protective protein	N333	1	4	Glycan degradation	
PRNP	PRIO_HUMAN	P04156	Major prion protein	N181	5	4	Unknown	Yes (74)
PTK7	PTK7_HUMAN	Q13308	Tyrosine-protein kinase-like 7	N405	5	1	Adhesion molecule	
PTK7	PTK7_HUMAN	Q13308	Tyrosine-protein kinase-like 7	N567	5	3	Adhesion molecule	
PVR	PVR_HUMAN	P15151	Poliovirus receptor	N120	4	3	Immune regulator	

Gene name	Entry name	Accession Number ^a	Glycoprotein	Glycosite(s)	Observed TN AAL ^b	Observed TN SNA ^b	Function ^c	Known Fucosylation or Sialylation ^d
SEZ6L2	SE6L2_HUMAN	Q6UXD5	Seizure 6-like protein 2	N247	3	2		
SEZ6L2	SE6L2_HUMAN	Q6UXD5	Seizure 6-like protein 2	N373	3	1	Unknown	
SPARC	SPRC_HUMAN	P09486	SPARC (Osteonectin)	N116	5	3	Cell growth, Differentiation	Yes (75)
SUSD5	SUSD5_HUMAN	O60279	Sushi domain-containing protein 5	N354	4	4	Unknown	
ABI3BP	TARSH_HUMAN	Q7Z7G0	Target of Nesh-SH3	N44	3	2	Cell migration	
TFPI	TFPI1_HUMAN	P10646	Tissue factor pathway inhibitor	N145	5	4	Complement cascade	Yes (76)
TGFB1	TGFB1_HUMAN	P01137	Transforming growth factor beta-1	N82	4	1	TGF-beta pathway	Yes (77)
TGFB2	TGFB2_HUMAN	P61812	Transforming growth factor beta-2	N241	5	4	TGF-beta pathway	
THBS3	TSP3_HUMAN	P49746	Thrombospondin-3	N407	3	2	Adhesion molecule	
TWSG1	TWSG1_HUMAN	Q9GZX9	Twisted gastrulation protein homolog 1	N52	3	2	Cell growth, Differentiation	
TXNDC15	TXD15_HUMAN	Q96J42	Thioredoxin domain-containing protein 15	N293	3	2	Unknown	
AXL	UFO_HUMAN	P30530	Tyrosine-protein kinase receptor UFO	N43	4	4	Receptor	
AXL	UFO_HUMAN	P30530	Tyrosine-protein kinase receptor UFO	N157	4	2	Receptor	
AXL	UFO_HUMAN	P30530	Tyrosine-protein kinase receptor UFO	N198	4	2	Receptor	
AXL	UFO_HUMAN	P30530	Tyrosine-protein kinase receptor UFO	N339	4	3	Receptor	
PLAUR	UPAR_HUMAN	Q03405	Urokinase plasminogen activator surface receptor	N222	2	3	ECM remodeling	Yes (78)
PLAU	UROK_HUMAN	P00749	Urokinase-type plasminogen activator	N322	4	4	ECM remodeling	
Not available	YK047_HUMAN	Q68D85	Putative Ig-like domain-containing protein	N242	3	0	Unknown	

^aUniprot accession number.

^bNumber of triple-negative (TN) cell lines in which a glycosite was observed with this lectin.

^cUniprot database annotation.

^dReferences in parentheses.

^eDenotes glycoproteins observed in healthy plasma following AAL or SNA enrichment (24).