



Published in final edited form as:

J Proteomics. 2014 January 16; 96: 173–183. doi:10.1016/j.jprot.2013.11.006.

Using a Cell Line Breast Cancer Progression System to Identify Biomarker Candidates

Ten-Yang Yen, Nicole Haste, Leslie C. Timpe, Christina Litsakos-Cheung, Roger Yen, and Bruce A. Macher*

Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, CA 94132

Abstract

Secreted and plasma membrane glycoproteins are considered excellent candidates for disease biomarkers. Herein we describe the identification of secreted and plasma membrane glycoproteins that are differentially expressed among a family of three breast cancer cell lines that models the progression of breast cancer. Using two-dimensional liquid chromatography-tandem mass spectrometry we identified more than 40 glycoproteins that were differentially expressed in either the premalignant (MCF10AT) or the fully malignant (MCF10CA1a) cell lines of this model system. Comparative analysis revealed that the differentially expressed breast cancer progression-associated glycoproteins were among the most highly expressed in the malignant (MCF10CA1a) breast cancer cell line; a subset of these were detected only in the malignant line; and others were detected in the malignant line at levels 25 to 50 times greater than in the benign (MCF10A) line. Using the results from this model cell system as a guide, we then carried out glycoproteomic analyses of normal and cancerous breast tissue lysates. Eleven of the glycoproteins differentially expressed in the breast cell lines were identified in the tissue lysates. Among these glycoproteins, collagen alpha-1 (XII) chain was expressed at dramatically higher (~10-fold) levels in breast cancer than in normal tissue.

Keywords

Breast cancer; glycoproteomics; differential expression; non-malignant vs. malignant; label-free quantitation; spectral counts

Introduction

Breast cancer is a slowly developing disease that involves the accumulation of genomic aberrations including amplifications, deletions and rearrangements. These alterations occur in genes associated with the growth, differentiation and death of cells and are critical for tumor development and metastasis.¹ Although much is known about the molecular changes

© 2013 Elsevier B.V. All rights reserved.

*Corresponding Author: Dr. Bruce A. Macher, Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, CA 94132, Telephone: 415-338-6078, FAX: 415-338-6253, macher@sfsu.edu.

Supporting Information Available: Supporting Information, including protein and peptide identifications are included in the Supplementary Tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

associated with the development of breast cancer, the events leading to cellular transformation and a metastatic phenotype have not been fully identified.

To elucidate the steps in breast cancer progression, model systems that reflect stages of breast cancer initiation and progression have been developed. One such model is based on the insertion of an activated Ha-Ras oncogene into a spontaneously immortalized human breast epithelial cell line, MCF10A², derived from a woman with fibrocystic disease. The resulting cell line, MCF10AT, has many characteristics of a malignant phenotype including anchorage-independent growth, growth in the absence of hormones or growth factors, and limited tumorigenicity in immune deficient mice, and is therefore referred to as a premalignant cell line.³ Additional cell lines, including MCF10CA1a with fully malignant characteristics, have been developed from xenografts isolated from immune deficient mice that were injected with the MCF10AT cell line. Our laboratory has been using this series of MCF10A cell lines--MCF10A (a benign line); MCF10AT (a premalignant line); and MCF10CA1a (a fully malignant line)--as a model system to investigate the changes that occur during breast cancer progression.

To achieve an invasive phenotype, malignant cells must alter many of their biochemical properties. These alterations often involve proteins that participate in cell-cell and cell-extracellular interactions which are typically mediated by glycoproteins.⁴⁻⁵ Characterization of the changes that occur in plasma membrane and secreted glycoproteins is a critical next step in developing a means of monitoring tumor progression and identifying therapeutic targets.

In the current study we have employed a well-established approach to identify cell surface and secreted glycoproteins from three MCF10A cell lines that constitute a model system of breast cancer progression. The glycoprotein profiles of each of the three cell lines reveal alterations in the expression levels of cell surface and secreted proteins associated with each stage in this progression series. After generating a list of candidates based on these profiles, we conducted a comparative analysis of these differentially expressed glycoproteins in normal breast tissue and breast cancer tissue. Our results demonstrate that collagen alpha-1 (XII) chain is a potential biomarker for breast cancer.

Experimental Methods

Cell Culture

Three breast cell lines -- a benign breast tumor cell line (MCF10A), and two derivatives (MCF10AT; MCF10CA1a) -- were grown in 10 cm culture dishes with 10ml of culture medium. The growth medium was composed of 1:1 DMEM: Ham's F12 with L-glutamine and 15mM HEPES (Invitrogen), 5% NZ horse serum (Gibco), insulin (10ug/ml, Sigma), human EGF (20ng/ml, Peprotech), hydrocortisone (0.5ug/ml, Sigma), and cholera toxin (0.1ug/ml, Sigma). MCF10CA1a was cultured in medium containing 1:1 DMEM: Ham's F12 with L-glutamine and 15mM HEPES (Invitrogen), 5% NZ Horse Serum (Gibco), and 1% antibiotics (100 units of Pen and 100 units of Strep, Hyclone). All three cell lines were grown at 37°C with 5% CO₂. Each sample (12ml lysate) for analysis was prepared from 15 culture dishes, yielding 1.17±0.58 mg of total protein.

Periodate Oxidation

Intact cells were treated with periodate to oxidize monosaccharides within the carbohydrate chains of secreted and cell surface glycoproteins. These glycoproteins were then enriched using hydrazide modified magnetic beads and identified by LC/MS/MS analyses as previously described.⁶⁻⁸ Briefly, the cell lines were grown to ~90% confluence and cells were oxidized with 10mM NaIO₄ in the dark at 25°C for 1 hour, after which they were lysed

with a pH 5 sodium acetate buffer of 1% octyl- β -D-1-thioglucopyranoside and 1% protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO). The cell residue was scraped from the dishes and homogenized by multiple passes through a syringe with a series of different needle sizes ranging from 19 to 27½ gauge. The lysates were clarified by centrifugation at 14,000 rpm for 8 minutes at 4°C, and the supernatant collected. Total protein content was determined using the Bradford assay, and the supernatant was frozen at -20°C until the samples were enriched for glycoproteins using hydrazide magnetic beads.

Glycoprotein Enrichment Reduction, Alkylation, Denaturation and Trypsin Digestion

Cell lysates were spiked with 5µg of periodate oxidized chicken ovalbumin (Sigma-Aldrich, St. Louis, MO) and coupled to hydrazide modified magnetic beads (30mg, 1µm size, Bioclone, San Diego, CA) using an Eppendorf thermomixer at 700 rpm at 25°C for 12-18h. The beads were washed twice with NaCl (1.5M), 50% methanol/water, 50% acetonitrile/water and 8M urea. Proteins were reduced with 50mM dithiothreitol for 40 minutes at 40°C, and alkylated with 1ml of 50mM iodoacetamide for 30 minutes at 25°C in the dark. The glycoproteins were digested with 0.5ml of trypsin (20ng/µl; Promega, Madison, WI) in 50mM ammonium bicarbonate buffer at 37°C with constant mixing for 12h. After digestion, the tryptic fraction was collected, and the hydrazide magnetic beads were washed with 3ml of 50mM ammonium bicarbonate to collect any remaining tryptic peptides. The combined eluate was processed by solid phase extraction (SPE), dried using a Speed-Vac apparatus, and stored at 4°C prior to mass spectrometric analysis.

N-Glycopeptide Release from the Hydrazide Resin with N-Glycosidase F

N The N-linked glycopeptides bound to the hydrazide magnetic beads were released from the beads with N-glycosidase F (PNGase F, Glyco N-Glycanase 200mU diluted 4-fold, Prozyme) (4µl in 0.5ml of 50mM ammonium bicarbonate buffer) by incubating the solution overnight at 37°C with constant mixing. O¹⁸ water was included in the PNGaseF reaction—resulting in the incorporation of O¹⁸ rather than of O¹⁶ during hydrolysis—to increase the level of confidence in glycopeptide identification.⁹⁻¹⁰

The released N-linked glycopeptide fraction was collected and the beads were rinsed with 3ml of 50mM ammonium bicarbonate buffer. The ammonium bicarbonate rinse solution was collected and combined with the PNGase F released fraction, processed by SPE, dried using a Speed-Vac apparatus, and stored at 4°C prior to mass spectrometric analysis.

Peptide/Protein Identifications by ESI-MS/MS Analysis

Tryptic peptides and the deglycosylated N-linked peptides derived from each cell lysate were separately analyzed using an extensive set of LC/MS/MS conditions to maximize glycoprotein identification and to improve the reproducible detection of low abundance glycoproteins. The tryptic peptides and the PNGase F treated N-linked glycopeptides were analyzed by liquid chromatography/electrospray ionization-tandem mass spectrometry (LC/ESI-MS/MS) using a Thermo LTQ ion trap mass spectrometer with dual Thermo Surveyor HPLC pump systems (Thermo Fisher, San Jose, CA) for varying periods (30, 45, 60 and 90s) of dynamic exclusion (DE) and using three different gas phase fractionation settings. LC/ESI-MS/MS analyses were conducted using a C18 column (75µm × 130mm). In the reverse phase chromatography, a solution of 0.1% HCOOH in water was used for mobile phase A and 0.1% HCOOH/acetonitrile for mobile phase B. A four-step, linear gradient was used for nanoLC separation (5% to 35% B in the first 65 min; 35% to 80% B in the next 10 min; held at 80% B for 5 min, and returned to 5% B during the final 10 min). The ESI-MS/MS data acquisition software was set to collect ion signals from the eluted peptides using an automatic, data-dependent scan procedure with top 4 triple plays. The full scan mass range was set from m/z 400 to 1800. Additionally, on-line 2D-LC/ESI-MS/MS analyses were

conducted to analyze the trypsin-digested fraction. Seven fractions were eluted from a SCX column (Thermo Fisher, BioBasic SCX column, 320 μ m \times 100mm) using various concentrations of NH₄Cl (0, 20, 40, 60, 80, 200, 400mM, and a 2nd wash 400mM). The resulting fractions were desalted in-line with dual C18 trap columns (300 μ m \times 5mm, Agilent); each fraction was chromatographically resolved by 1D-LC with DE=1 at 45s and the LC elution time increased 120min. Two algorithms (Mascot v2.3¹¹ and X!Tandem 2007.01.01.1¹²) were employed to identify peptides from the resulting MS/MS spectra by searching against the combined human protein database (total 22673 proteins) extracted from SwissProt (v57.14; 2010 February) using taxonomy "homo sapiens" (22670 proteins). Search parameters were set as follows: parent and fragment ion tolerances of 1.6 and 0.8 Da, respectively; carbamidomethyl (+57 Da) modification of Cys as a fixed modification; deamidation (+1 Da) of Asn or Gln, and oxidation of Met as variable modifications; and trypsin as the protease with a maximum of 2 missed cleavages. Scaffold (Proteome Software) was used to merge and summarize the data obtained from the twenty-four runs of LC/MS/MS protein identification analyses for each cell lysate preparation (8 \times 1D-LC/MS/MS acquisition for the PNGaseF released sample; 8 \times 1D-LC/MS/MS acquisition and 8 \times 2D-LC/MS/MS acquisition for the tryptic digested sample). Protein identifications were based on a minimum detection of 2 peptides with 99% protein identification probability using the algorithm ProteinProphet.¹³ Each peptide had a minimum peptide identification probability of 95% using the algorithm PeptideProphet.¹⁴ Peptides identified from four biological replicates of each of the three MCF10A cell lines are listed in Supplemental Table 1. The average false positive rate for peptide identification in this study was 2.65 \pm 2.42%, based on results obtained with PeptideProphet. ProteinID Finder (Proteome Solutions) was used to determine whether the peptide was derived from a glycoprotein and to extract protein information (e.g., subcellular location) from the UniProt database for each identified protein.

Breast Tissue Lysates

Lysates (100 micrograms of protein) from normal and cancerous breast tissue were obtained from Origene Technologies (Rockville, MD). The glycoproteins within the lysates were oxidized with 20mM NaIO₄ in the dark at 25°C for 1 hour. Oxidation was terminated by adding sodium sulfite at a final concentration of 40mM. Glycoprotein enrichment was carried out by the procedures described above.

The glycoprotein fraction from each lysate was analyzed by LC/ESI-MS/MS using a hybrid quadrupole orbitrap mass analyzer (Thermo QExactive, Bremen, Germany) coupled with a Thermo-Dionex HPLC system (Germering, Germany). LC/MS/MS analysis was employed using a data-dependent acquisition with a cycle of MS/MS on the 10 most abundant precursors. The mass resolution was set as 70,000 at m/z 200 for the full scan (400-1,800 Th), and 17,500 or 35,000 for the MS/MS scan, respectively. The maximum ion accumulation time was 100s and the dynamic exclusion was 10s. The HCD collision energy was set at 27%. Protein identification was carried out using Proteome Discoverer 1.4 with Sequest HT at a false discovery rate less than 1%.

Statistical Analyses

Significance analysis for microarrays (SAM)¹⁵ was performed using the MultiExperimentViewer (MeV, v4.7) program.¹⁶ A value of 0.5 spectral counts was added to avoid the log₂ calculation problem derived from a zero value for glycoproteins that were undetected in some samples.

Western blot analysis

Five or 10 μ g of cell lysate was loaded per lane on a 4-12% NuPAGE bis-Tris gel (Invitrogen) under reducing (30mM beta-mercaptoethanol) or non-reducing conditions. A voltage of 200V was applied to the gel for 35 min. The proteins in the gel were transferred to a nitrocellulose membrane using the iBlot Dry Blotting System (Invitrogen). The membrane was blocked with 0.01% BSA in PBS and then incubated with one of the following primary antibodies: rabbit anti-tissue factor (1:500 dilution of Genetex antibody #GTX100808), chicken anti-CEACAM5 (1:2000 dilution of ProSci, Inc. antibody XW-8078), goat anti-lipocalin 2 (1:1000 dilution of ProSci, Inc. antibody 45-838) or mouse anti-cathepsin D (1:500, Pro-Sci Inc. 48-051). After washing with TBS containing 1% Tween 20, the blots were incubated with the appropriate secondary antibodies: anti-rabbit-alkaline phosphatase (1:500), anti-chicken-alkaline phosphatase (1:1000), anti-goat-alkaline phosphatase (1:500), or anti-mouse-alkaline phosphatase (1:500), respectively. Detection was achieved using the nitro blue tetrazolium chloride/5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt alkaline phosphatase substrate reaction.

Western blotting was also used to analyze breast tissue lysates for the expression of collagen alpha-1(XII) chain using a mouse monoclonal antibody (sc-166020, Santa Cruz Biotechnology, Santa Cruz, CA) at a 1:500 dilution in Li-Cor Blocking Buffer (Li-Cor Biosciences) with 0.1% Tween-20. The membrane was then treated with IRDye CW 800 Anti-Mouse (926-32212, Li-Cor Biosciences) antibody at a 1:15,000 dilution. The blot was imaged using the Odyssey CLx (Li-Cor Biosciences, Lincoln, NE) infrared scanner and viewed using ImageStudio software (Li-Cor Biosciences).

Results and Discussion

Breast cancer progression-related alterations in cell surface and secreted glycoprotein expression profiles

Cell surface and secreted glycoproteins from the three MCF10A cell lines were identified using a well-established approach: periodate oxidation of intact cells and covalent capture of the glycoproteins via their oxidized glycan chains.¹⁷⁻¹⁸ Tryptic peptides and the deglycosylated N-linked peptides derived from each cell lysate were separately analyzed using a set of 24 LC/MS/MS conditions to maximize glycoprotein identification and to improve the reproducible detection of low abundance glycoproteins.¹⁹⁻²¹

We analyzed four biological replicates of each cell line. The combination of tryptic peptides and PNGase F released N-linked glycopeptides, which resulted in the identification of 363 glycoproteins exclusive of HLA antigen sequences (Supplemental Table 2). On average, 259 glycoproteins were detected per cell line. The relative quantity of each protein was measured using a label free method to calculate the total number of matched MS/MS spectra (spectral counts) for the tryptic- and PNGase F-released peptides associated with each identified glycoprotein. As previously established, comparison of the spectral counts for one glycoprotein across each of the different samples provides a reliable estimate of their relative abundances.²²⁻²⁴ At least 60% of the glycoproteins observed with more than 10 spectral counts were identified in all 4 biological replicates for each cell line, and more than 80% of these glycoproteins were identified in 3 out of 4 of the replicates. Given that cell lines MCF10AT and MCF10CA1a are derived from MCF10A, significant overlap among the identified glycoproteins in this family of cell lines is not surprising. Additionally, as shown in Figure 1, the correlation between the identified glycoproteins among the cell lines in the MCF10A family is significantly greater than the correlation between the glycoproteins from any cell lines in this family and those from an unrelated basal breast cell line (e.g. HCC1143).

Differential expression of glycoproteins in the tumor progression model

Although the glycoproteomic profiles of the three MCF10A cell lines (MCF10A, MCF10AT, MCF10CA1a) are qualitatively similar, statistically significant differences were detected among the glycoproteins identified from these three lines. Significance analysis for microarrays (SAM)¹⁵ was used to determine which glycoproteins were differentially expressed by first calculating a modified *t* statistic for the difference in the means of the groups for each glycoprotein and then using a permutation test to determine significance levels.

Comparison of the glycoprotein profiles of the benign (MCF10A) vs. the malignant (MCF10CA1a) cell lines revealed 38 differentially expressed glycoproteins (with a false discovery rate ~ 0.0006). Of these, 33 were more highly expressed in the malignant line, and 5 were more highly expressed in the benign line. Further, of the 33 that were more highly expressed in the malignant line, 12 were undetected in the benign line; and none of the 5 that were more highly expressed in the benign line were detected in the malignant line (see Figure 2).

Many of the glycoproteins differentially expressed in the malignant line (MCF10CA1a) were also among the glycoproteins most highly expressed in this line, with spectral counts putting them within the top 8-20% of all glycoproteins identified. Tissue factor (gene name F3) has the highest level of expression, with more than 130 spectral counts, and its identification is based on a sequence coverage of 71%.

To independently verify that tissue factor is differentially expressed in the MCF10A cell lines, cell lysates from the three lines were analyzed with an anti-tissue factor antibody by Western blotting. As illustrated by Figure 3, no bands for tissue factor appear in the lane containing lysate from MCF10A, whereas the lane containing lysate from MCF10CA1a has a major band migrating at ~ 50 kD. These results independently confirm the results of the mass spectrometry.

We also carried out Western blot analyses with commercially available antibodies against two other glycoproteins--carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5), and neutrophil gelatinase-associated lipocalin 2 (LCN2)--that are among the proteins which were significantly overexpressed in the malignant (MCF10CA1a) line. In each case, results obtained by Western blotting substantiated the mass spectrometry results (Figure 3). Cathepsin D, expressed at significant levels but varying levels (MCF10A < MCF10AT < MCF10CA1a) in each of the cell lines, is shown for comparison.

Glycoprotein expression change associated with stages of tumor progression

For some glycoproteins, the change in expression level was greatest in the first step of our model of breast cancer progression, from benign (MCF10A) to pre-malignant (MCF10AT). In others, the greatest change occurred in the second step, from pre-malignant (MCF10AT) to fully malignant (MCF10CA1a). Table 1 compares the level of each of the 38 identified glycoproteins (Figure 1) across the three breast cell lines. Among these 38 glycoproteins, 3 (CEACAM5, NGFR, CD164) were detected exclusively in MCF10CA1a, suggesting that their expression correlates with a fully malignant cell phenotype. Several additional glycoproteins that were detected at high levels in MCF10CA1a were also found in MCF10A and MCF10AT at the lower limit of detection (1-2 spectral counts), suggesting that their expression also correlates with a fully malignant phenotype. Further substantiating this assignment, four glycoproteins (F3, ECE1, CEACAM1, MF12) associated with the fully malignant phenotype were detected in MCF10CA1a at levels 25 to 50-fold greater than in MCF10A, the benign cell line.

As shown in Table 1 (Higher Expression in MCF10AT and MCF10CA1a), SAM analysis revealed that 5 (LYPD3, ITGB6, SPINT1, CD97 and PROM2) of the 38 differentially expressed glycoproteins were expressed at dramatically higher levels in the premalignant (MCF10AT) than in the benign (MCF10A) stage. In contrast, these 5 glycoproteins were expressed at only a slightly higher level (an increase of less than 2fold) in the fully malignant (MCF10CA1a) than in the premalignant (MCF10AT) stage. We conclude, therefore, that the expression of these 5 glycoproteins is primarily correlated with RAS transfection-mediated changes that convert the benign line into a premalignant line.

Differential expression of glycoproteins associated with premalignant to malignant transition, MCF10AT vs. MCF10CA1a

In addition to the glycoproteins listed in Table 1, six additional glycoproteins (labeled with an asterisk in Figure 4) were determined by SAM analysis to be differentially expressed at significant levels in the premalignant cell line MCF10AT vs. the malignant cell line MCF10CA1a. These six glycoproteins include three protease inhibitors: metalloproteinase inhibitor 1 (TIMP1), alpha-1-chymotrypsin, and alpha-2-macroglobulin-like protein 1. Our results show that TIMP1 and alpha-1-chymotrypsin are expressed at significantly higher levels in the malignant cell line, whereas alpha-2-macroglobulin-like protein 1 is expressed at significantly lower levels in the malignant line. Although proteases are often associated with cancer and particularly with invasiveness, studies have also found associations between some protease inhibitors, including TIMP1, and cancer. In fact, higher levels of TIMP1 have been shown to predict poor outcomes in breast cancer (reviewed in ²⁵). While no prior studies have demonstrated an association between alpha-1-chymotrypsin and alpha-2-macroglobulin-like protein 1, and breast cancer, limited data have shown a relationship between steroid hormones and secretion of alpha-1-chymotrypsin.²⁶

Our results indicate that mucin-1 expression was significantly reduced in the transition from the premalignant (MCF10AT) to the malignant state (MCF10CA1a). This finding contradicts prior reports that mucin-1 is overexpressed in many cancers including breast cancer.²⁷⁻²⁸ Other studies²⁹ in animal models report variable levels of mucin-1 expression in breast cancer cell lines without correlation to the potential malignancy of these lines. In contrast to mucin-1, we found that sialomucin (CD164) was significantly overexpressed in the malignant (MCF10CA1a) compared to the pre-malignant (MCF10AT) cell lines in our model system.

Comparison with other work on glycoproteins of the MCF10A family of cell lines

Mbeunkui et al.³⁰ evaluated the secretome of four MCF10A derived cell lines, including two (MCF10A and MCF10AT) reported on here. They reported the identification of more than 200 proteins and provided information on ~40 proteins, of which less than half are glycoproteins. Among the glycoproteins they reported, we identified only one, lipocalin 2, as being significantly differentially expressed among the MCF10A cell lines; it was most highly expressed in the malignant cell line MCF10CA1a. Mbeunkui et al.³⁰ did not detect lipocalin 2 in either MCF10A or MCF10AT; however, they did identify it in the tumorigenic, locally-invasive cell line MCF10DCIS.

Wang et al.³¹ used the MCF10AT and MCF10CA1a cell lines as protein sources in a study evaluating the comparative effectiveness of two cell lysate approaches for isolating membrane proteins. These investigators utilized two (ConA and WGA) lectin chromatography columns to enrich for glycoproteins that were posttranslationally modified with different N-linked glycans. In comparing the two MCF10A cell lines, Wang et al.³¹ reported the identification of a number of glycoproteins that were differentially expressed, defined as a two-fold change.

There was no correlation between the list of glycoproteins reported by Wang et al.³¹ and the glycoproteins identified as being differentially expressed in the present study. Since the approaches used in these two studies differ substantially, the lack of correlation between the results is unsurprising. For example, Wang et al.³¹ identified and quantified only glycoproteins that contained a glycan bound by one of the two lectins used, and only that portion of each glycoprotein that contained a glycan bound by the lectins. We identified and quantified only glycoproteins that are secreted or found in the plasma membrane. Further, most of the glycoproteins we identified were not part of the set reported by Wang et al.³¹; our glycoprotein set was larger than their reported set by a factor of eight. Each analytical approach produced a unique set of differentially expressed glycoproteins. We specifically targeted cell surface and secreted glycoproteins because these are the proteins most likely to be of diagnostic and therapeutic value.

Recently, Boersema et al.³² employed a lectin (ConA and WGA used together) capture technique to obtain glycoprotein profiles from conditioned media generated by primary cultures of normal human mammary epithelia and cell lines derived from benign (including MCF10A) and malignant human breast tumors. Based on an hierarchical cluster analysis of the N-linked glycopeptides they identified and quantified, Boersema et al.³² observed that the cell lines belonging to the same cancer stage (II, III or metastatic) were grouped together. Clusters of glycoproteins that trended significantly up or down in correlation with cancer stage were identified as potential breast cancer markers. Although our study differed from that of Boersema et al.³² in methodology, majority of cell lines used, and approaches to data analyses, both studies identified CEACAM1 as a breast cancer associated marker.

We then reanalyzed the expression data from Boersema et al.³² (their supplemental Table 4) for glycoproteins that were differentially expressed between their malignant and non-malignant cell lines. Of the five glycoproteins identified that fit this criteria, four (CEACAM5, LCN2, GOLM1, SEZ6L2) showed significantly increased expression among their malignant cell lines, and one (CPVL) showed significantly decreased expression among their malignant lines, consistent with the results of our study. The consistency of these observations across two independent studies of breast cancer progression strongly suggests that this glycoprotein set contains promising markers of breast cancer.

Finally, Lai et al.³³ compared the secretome (conditioned medium) and total lysates of MCF10A with that of two breast cancer lines (MCF7 and MB-MDA-231) using two dimensional differential gel electrophoresis and MALDI-TOF mass spectrometry. Of the proteins that were reported by these authors to be secreted and differentially expressed among the three cell lines only five are listed as glycoproteins in the UniProt database. Ten additional glycoproteins were reported in the total lysates from the three cell lines. Ten of the glycoproteins reported in the Lai et al. study were also identified in our study. However, none of these glycoproteins were found to be differentially expressed among the MCF10A family of cell lines. This variation in outcomes is unsurprising given the distinctive differences between our experimental design and that of Lai's group. In fact, these findings underscore the benefits of applying a variety of experimental designs and proteomic analyses in the search for potential biomarker candidates.

Many of the glycoproteins associated with the malignant phenotype are important in breast cancer and tumor aggressiveness

A review of the literature shows that 17 of the 22 glycoproteins differentially expressed in MCF10CA1a have been implicated as important in breast cancer. At least five of the glycoproteins in this group have also been associated with cancer progression (Table 1). Several studies (reviewed in van den Berg et al.³⁴) have demonstrated that tissue factor is highly expressed in solid tumors including breast tumors and that tissue factor levels are

correlated with tumor aggressiveness. For example, in a study using immunohistochemistry to examine breast tumors from more than 200 women, Ueno et al.³⁵ found tissue factor expression to be associated with breast tumors in more than 90 percent of the samples analyzed and, further, that patients with tumor associated tissue factor expression had poorer prognoses than those of patients with tissue factor negative tumors. In addition, Versteeg et al.³⁶ have shown that tissue factor is associated with beta 1 integrin in aggressive breast cancer cells. Knockdown of beta 1 integrin via RNA interference has been found to reduce tumor take, growth and angiogenesis in triple-negative breast cancer.³⁷

Endothelin-converting enzyme 1 is another protein that is expressed at dramatically higher levels in the MCF10CA1a cell line than in MCF10A cell line. Endothelin-converting enzyme 1 and neprilysin regulate the level of endothelin-1, which affects cells through two G coupled receptors. Activation of these receptors has been reported to result in tumor invasion and to increase proliferation, angiogenesis and inhibition of apoptosis. Our results demonstrate that whereas MCF10CA1a cells have much higher levels of endothelin-converting enzyme 1 than MCF10A cells, neprilysin is expressed only in MCF10A cells (see bottom of Table 1). The expression levels of endothelin-converting enzyme 1 and neprilysin in MCF10CA1a cells would be expected to result in high levels of expression of endothelin-1. A study by Smollich et al.³⁸ evaluating the levels of endothelin-converting enzyme 1 and neprilysin in 600 breast cancer tissue samples showed that patients with tumors in which endothelin-converting enzyme 1 was over expressed had more frequent disease recurrence, whereas patients with tumors in which neprilysin was over expressed experienced less frequent metastasis and higher rates of disease-free survival. Thus, findings from our work using the MCF10A family of cell lines as a model system match *in vivo* observations of changes in endothelin-converting enzyme 1 and neprilysin that occur in breast cancer.

As shown in Table 1, most of the glycoproteins that are differentially expressed in the MCF10A family of cell lines have biological properties that are hallmarks of cancer¹ and many of these glycoproteins--approximately two-thirds--have been directly associated with breast cancer. In some instances we found no information in the published literature substantiating a specific association between breast cancer and the glycoproteins we identified in Table 1, but abundant evidence associating these glycoproteins with other cancers (e.g., seizure 6-like protein 2 associated with lung cancer³⁹). Our results thus provide new information on the relationship of several glycoproteins and breast cancer. It is important to point out that while the majority of changes we observed in glycoprotein expression within the MCF10A family of cell lines are correlated with previous observations in the literature, some of our findings contrast with those previously reported. For example, Schroder et al.⁴⁰ demonstrated that ICAM-1 expression is associated with a more aggressive breast tumor phenotype, whereas we detected ICAM1 in only the benign (MCF10A) cell line. The absence of a correlation between ICAM1 levels in the MCF10A family of cell lines and those in breast tumors is understandable because cell lines do not experience the microenvironmental influences to which breast tumors are exposed. The relevance of the tumor microenvironment as a factor in breast tumor ICAM1 expression is supported by the findings of Ogawa et al,⁴¹ who showed that patients with ICAM1-positive tumors had better outcomes (i.e., fewer relapses and higher survival rates) than patients with ICAM-1 negative tumors. Ogawa et al. speculate that ICAM1 expression on tumor cells may have a role as a suppressor of tumor progression modulated by the host immune system.

Breast Cancer Tissues Expression of Glycoproteins with Altered Expression Levels in the MCF10A Model System

As described in preceding paragraphs, the expression level of several of the glycoproteins differentially expressed among the MCF10A cell lines have been evaluated in breast tissues via antibody-mediated studies. To further investigate the expression of these glycoproteins in breast tissue using our glycoproteomic approach, we analyzed small (100 micrograms total protein) quantities of lysates from normal and breast cancer tissues. LC/MS/MS analysis was employed using a high resolution hybrid quadrupole orbitrap mass analyzer to assure protein identification with the highest confidence. More than 400 glycoproteins (spectral counts 2 or more) were identified from tissue lysates (Supplemental Table 2). As shown in Table 2, we detected a total of 11 glycoproteins which we identified as differentially expressed among the MCF10A cell lines: nine glycoproteins that were more highly expressed in the malignant MCF10A lines (MCF10AT and MCF10ACA1a) and two that were more highly expressed in the benign (MCF10A) line. Of the nine glycoproteins that were more highly expressed in the malignant cell lines, six were identified only in tumor tissue lysates. A seventh (collagen alpha-1(XII) chain) was expressed at significantly higher (>10-fold on average) levels among all of the tumor lysates compared to lysates from normal breast tissue. These initial results show that for 7 glycoproteins, the results obtained from the glycoproteomic analyses of the MCF10A cell line model are consistent with those obtained with tissue samples. Of these seven glycoproteins, collagen alpha-1(XII) chain was consistently and highly expressed in the tumor lysates (see Supplemental Table 3 for a list of identified peptides). Although this glycoprotein does not appear to have been previously associated with breast cancer, there is evidence that it is over expressed in colon cancer⁴². Further work is required to identify the cellular source of collagen alpha-1(XII) chain in breast tumor tissue.

Collagen alpha-1(XII) chain is a large protein composed of >3,000 amino acids, thought to be glycosylated with both N- and O-linked glycans, that generates a highly complex banding pattern on Western blots⁴³. To examine the molecular weight distribution of this glycoprotein in lysates from normal and tumorous breast tissue, we conducted comparative Western blot analyses (Figure 5). The normal lysates (lanes 4 and 5) produced a distinctive pattern in which the band of maximum intensity occurs at ~250,000 kD with a series of lower molecular weight bands. In contrast, the tumor lysates had patterns containing highly stained bands that ran at molecular weights above 250 kD, plus all of the bands observed in the normal lysates. Thus, the spectral counts (shown in Table 2) demonstrate breast tumor lysates contained quantitatively (spectral counts as shown in Table 2) more collagen alpha-1(XII) chain than lysates from normal breast tissue and the molecular weight distribution of bands differed significantly between lysates from normal breast tissue and those from tumorous breast tissue.

Within the tissue lysates, we also detected two glycoproteins (ICAM1 and Mucin 1) that we had previously determined were more highly expressed in the benign than in the malignant cell lines. However, in the case of the tissue samples these two glycoproteins were only detected in lysates from breast tumor tissues. Thus, there was no correlation between the expression pattern of ICAM1 and Mucin 1 between that observed in the MCF10A cell lines and that seen in the breast tissue lysate samples. However, our glycoproteomic results with breast tissue lysates are consistent with findings reported by others (see references and discussion in preceding paragraphs) for the expression of these two glycoproteins in normal vs. tumor tissues.

In summary, using a homogeneous family of cell lines to model the progression of breast cancer, we have identified a potential list of candidate biomarkers for breast cancer and applied this list to guide a glycoproteomic investigation of normal and tumor breast tissue.

Our findings demonstrate that several of the glycoproteins differentially expressed in the MCF10A cell line system are present in human breast tissue, that several of the differentially expressed glycoproteins were found in breast tumor tissue, but not in normal breast tissue, and most importantly that collagen alpha-1(XII) chain was differentially expressed in human breast tumor tissue. Future analyses with an expanded sampling of human breast tissues will be necessary to examine collagen alpha-1(XII), as well as other glycoproteins that are differentially expressed in the MCF10A model system, as potential biomarkers of breast cancer progression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for this work was provided by grants from the National Institutes of Health, R15CA164929 and P20 MD000544, and the National Science Foundation, Grant CHEM-0619163 and CHE-1228656. The authors thank Dr. Jose Lopez for providing the MCF10AT and MCF10CA1a cell lines, and Barbara Ustanko for editorial assistance.

References

1. Hanahan D, Weinberg, Robert A. Hallmarks of Cancer: The Next Generation. *Cell*. 2011; 144(5): 646–674. [PubMed: 21376230]
2. Basolo F, Elliott J, Tait L, Chen XQ, Maloney T, Russo IH, Pauley R, Momiki S, Caamano J, Klein-Szanto AJP, Koszalka M, Russo J. Transformation of Human Breast Epithelial Cells by c-Ha-ras Oncogene. *Molecular Carcinogenesis*. 1991; 4(1):25–35. [PubMed: 2009132]
3. Santner S, Dawson P, Tait L, Soule H, Eliason J, Mohamed A, Wolman S, Heppner G, Miller F. Malignant MCF10CA1 Cell Lines Derived from Premalignant Human Breast Epithelial MCF10AT Cells. *Breast Cancer Res Treat*. 2001; 65(2):101–110. [PubMed: 11261825]
4. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation*. 2009; 119(6):1420–1428. [PubMed: 19487818]
5. Kim YJ, Varki A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj J*. 1997; 14(5):569–76. [PubMed: 9298689]
6. Zou Z, Ibisate M, Zhou Y, Aebersold R, Xia Y, Zhang H. Synthesis and evaluation of superparamagnetic silica particles for extraction of glycopeptides in the microtiter plate format. *Analytical chemistry*. 2008; 80(4):1228–34. [PubMed: 18197692]
7. Berven FS, Ahmad R, Clauser KR, Carr SA. Optimizing Performance of Glycopeptide Capture for Plasma Proteomics. *Journal of Proteome Research*. 2010; 9(4):1706–1715. [PubMed: 20235580]
8. Yen TY, Macher BA, McDonald CA, Alleyne-Chin C, Timpe LC. Glycoprotein profiles of human breast cells demonstrate a clear clustering of normal/benign versus malignant cell lines and basal versus luminal cell lines. *J Proteome Res*. 2012; 11(2):656–67. [PubMed: 22106898]
9. Kuster B, Mann M. ¹⁸O-Labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal Chem*. 1999; 71(7):1431–1440. [PubMed: 10204042]
10. Yen TY, Dutta SM, Litsakos-Cheung C, Corona A, Timpe LC, Macher BA. Overcoming challenges and opening new opportunities in glycoproteomics. *Biomolecules*. 2013; 3(2):270–286.
11. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20(18):3551–3567. [PubMed: 10612281]
12. Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Communications in Mass Spectrometry*. 2003; 17(20):2310–2316. [PubMed: 14558131]

13. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical chemistry*. 2003; 75(17):4646–4658. [PubMed: 14632076]
14. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical chemistry*. 2002; 74(20):5383–5392. [PubMed: 12403597]
15. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*. 2001; 98(9):5116–5121.
16. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003; 34(2):374–8. [PubMed: 12613259]
17. McDonald CA, Yang JY, Marathe V, Yen TY, Macher BA. Combining results from lectin affinity chromatography and glyco-capture approaches substantially improves the coverage of the glycoproteome. *Mol Cell Proteomics*. 2009; 8(2):287–301. [PubMed: 18923192]
18. 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]
19. Spahr CS, Davis MT, McGinley MD, Robinson JH, Bures EJ, Beierle J, Mort J, Courchesne PL, Chen K, Wahl RC, Yu W, Luethy R, Patterson SD. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. *Proteomics*. 2001; 1(1):93–107. [PubMed: 11680902]
20. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*. 2001; 73(23):5683–90. [PubMed: 11774908]
21. Zhang Y, Wen Z, Washburn MP, Florens L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal Chem*. 2009; 81(15):6317–26. [PubMed: 19586016]
22. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76(14):4193–201. [PubMed: 15253663]
23. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*. 2005; 4(10):1487–502. [PubMed: 15979981]
24. Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, Washburn MP. Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics*. 2008; 7(4):631–44. [PubMed: 18029349]
25. Wurtz SO, Schrohl AS, Sorensen NM, Lademann U, Christensen IJ, Mouridsen H, Brunner N. Tissue inhibitor of metalloproteinases-1 in breast cancer. *Endocr Relat Cancer*. 2005; 12(2):215–27. [PubMed: 15947098]
26. Massot O, Baskevitch PP, Capony F, Garcia M, Rochefort H. Estradiol increases the production of alpha 1-antichymotrypsin in MCF7 and T47D human breast cancer cell lines. *Mol Cell Endocrinol*. 1985; 42(3):207–14. [PubMed: 3899774]
27. Kufe DW. MUC1-C oncoprotein as a target in breast cancer: activation of signaling pathways and therapeutic approaches. *Oncogene*. 2012
28. Tarp MA, Sorensen AL, Mandel U, Paulsen H, Burchell J, Taylor-Papadimitriou J, Clausen H. Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology*. 2007; 17(2):197–209. [PubMed: 17050588]
29. Walsh MD, Luckie SM, Cummings MC, Antalis TM, McGuckin MA. Heterogeneity of MUC1 expression by human breast carcinoma cell lines in vivo and in vitro. *Breast Cancer Res Treat*. 1999; 58(3):255–66. [PubMed: 10718487]
30. Mbeunkui F, Metge BJ, Shevde LA, Pannell LK. Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. *J Proteome Res*. 2007; 6(8):2993–3002. [PubMed: 17608509]

31. 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]
32. Boersema PJ, Geiger T, Wisniewski JR, Mann M. Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics.* 2013; 12(1):158–71. [PubMed: 23090970]
33. Lai TC, Chou HC, Chen YW, Lee TR, Chan HT, Shen HH, Lee WT, Lin ST, Lu YC, Wu CL, Chan HL. Secretomic and proteomic analysis of potential breast cancer markers by two-dimensional differential gel electrophoresis. *J Proteome Res.* 2010; 9(3):1302–1322. [PubMed: 20052998]
34. van den Berg YW, Osanto S, Reitsma PH, Versteeg HH. The relationship between tissue factor and cancer progression: insights from bench and bedside. *Blood.* 2012; 119(4):924–32. [PubMed: 22065595]
35. Ueno T, Toi M, Koike M, Nakamura S, Tominaga T. Tissue factor expression in breast cancer tissues: its correlation with prognosis and plasma concentration. *Br J Cancer.* 2000; 83(2):164–170. [PubMed: 10901365]
36. Versteeg HH, Schaffner F, Kerver M, Petersen HH, Ahamed J, Felding-Habermann B, Takada Y, Mueller BM, Ruf W. Inhibition of tissue factor signaling suppresses tumor growth. *Blood.* 2008; 111(1):190–9. [PubMed: 17901245]
37. Bianchi-Smiraglia A, Paesante S, Bakin AV. Integrin beta5 contributes to the tumorigenic potential of breast cancer cells through the Src-FAK and MEK-ERK signaling pathways. *Oncogene.* 2012
38. Smollich M, Gotte M, Yip GW, Yong ES, Kersting C, Fischgrabe J, Radke I, Kiesel L, Wulfig P. On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. *Breast Cancer Res Treat.* 2007; 106(3):361–9. [PubMed: 17295044]
39. Ishikawa N, Daigo Y, Takano A, Taniwaki M, Kato T, Tanaka S, Yasui W, Takeshima Y, Inai K, Nishimura H, Tsuchiya E, Kohno N, Nakamura Y. Characterization of SEZ6L2 cell-surface protein as a novel prognostic marker for lung cancer. *Cancer Science.* 2006; 97(8):737–745. [PubMed: 16863507]
40. Schröder C, Witzel I, Müller V, Krenkel S, Wirtz R, Jänicke F, Schumacher U, Milde-Langosch K. Prognostic value of intercellular adhesion molecule (ICAM)-1 expression in breast cancer. *J Cancer Res Clin Oncol.* 2011; 137(8):1193–1201. [PubMed: 21590495]
41. Ogawa Y, Hirakawa K, Nakata B, Fujihara T, Sawada T, Kato Y, Yoshikawa K, Sowa M. Expression of intercellular adhesion molecule-1 in invasive breast cancer reflects low growth potential, negative lymph node involvement, and good prognosis. *Clinical Cancer Research.* 1998; 4(1):31–36. [PubMed: 9516949]
42. Karagiannis GS, Petraki C, Prassas I, Saraon P, Musrap N, Dimitromanolakis A, Diamandis EP. Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget.* 2012; 3(3):267–85. [PubMed: 22408128]
43. Didangelos A, Yin X, Mandal K, Saje A, Smith A, Xu Q, Jahangiri M, Mayr M. Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. *Mol Cell Proteomics.* 2011; 10(8) M111.008128.

Abbreviations

EGF	epidermal growth factor
LC/ESI-MS/MS	liquid chromatography/electrospray ionization-tandem mass spectrometry
DE	dynamic exclusion
PNGase F	Peptide-N-Glycosidase F
SPE	solid phase extraction

SAM statistical analysis for microarrays
CD cluster of differentiation

Biological Significance

Identifying glycoproteins differentially expressed during cancer progression results in information on the biological processes and key pathways associated with cancer. In addition, new hypotheses and potential biomarkers result from these glycoproteomic studies. Our glycoproteomic analysis of this model of breast cancer provides a roadmap for future experimental interventions to further tease apart critical components of tumor progression

Highlights

Cell lines modeling breast cancer progression were glycoproteomically characterized.

Differentially expressed glycoproteins have properties that are hallmarks of cancer.

Two-thirds of the glycoproteins have been directly associated with breast cancer.

Sets of the glycoproteins are altered in premalignant and fully malignant stages.

This cell line system offers a viable model of breast cancer progression.

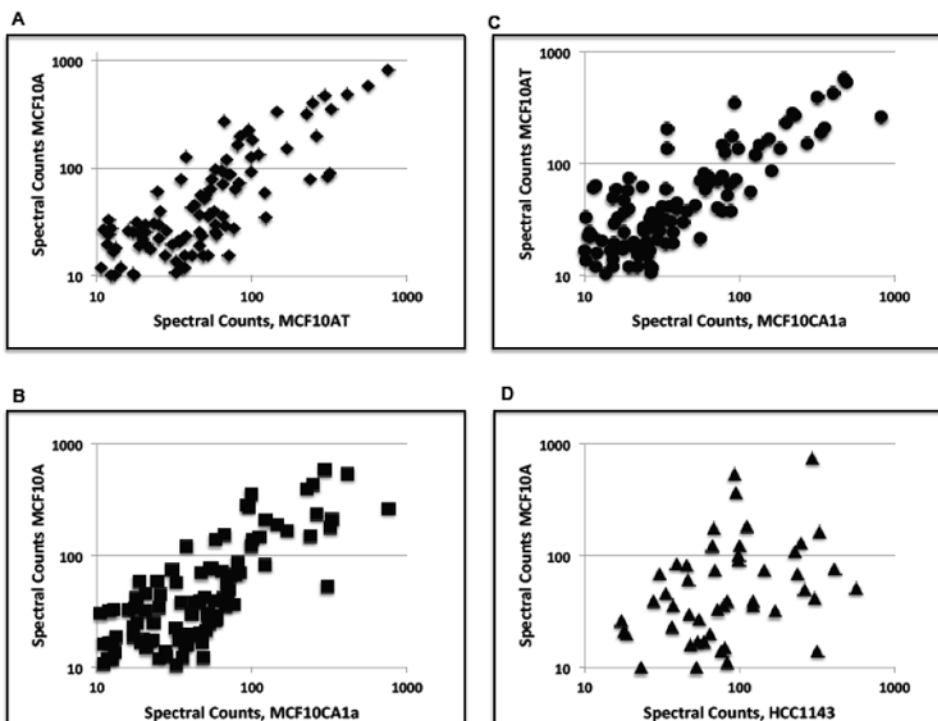


Figure 1. Comparisons of glycoprotein profiles of breast cancer cell lines. Glycoprotein profiles from MCF10A and its derivatives MCF10AT and MCF10CA1a show a greater correlation than the correlation between MCF10A and the breast cancer cell line HCC1143. Plots are log₁₀ spectral counts. (A) MCF10A vs. MCF10AT, $R^2 = 0.810$; (B) MCF10A vs. MCF10CA1a, $R^2 = 0.583$; (C) MCF10AT vs. MCF10CA1a, $R^2 = 0.667$; (D) MCF10A vs. HCC1143, $R^2 = 0.097$.

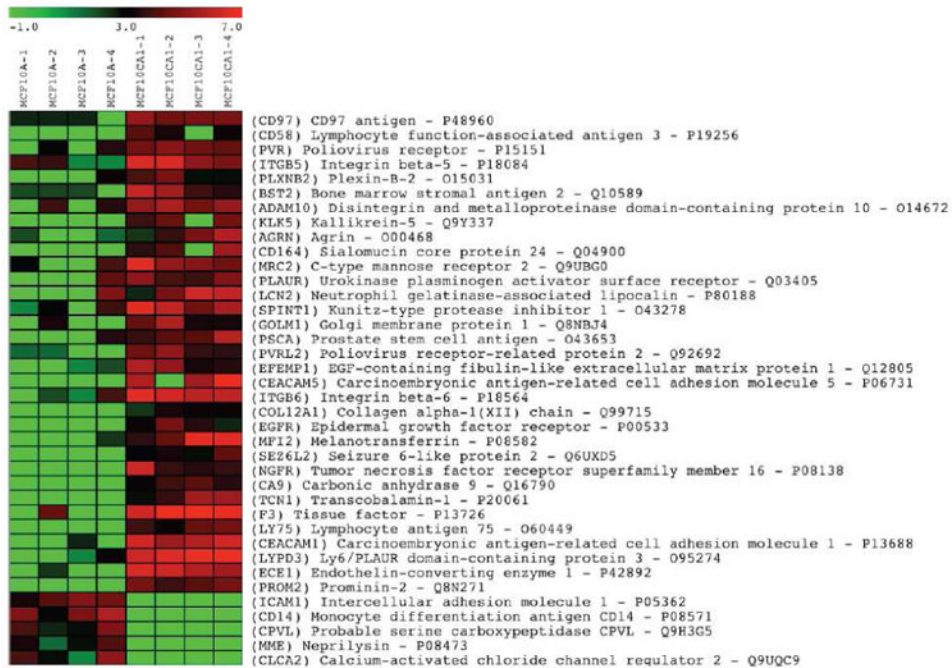


Figure 2. Differentially expressed glycoproteins comparing glycoprotein profiles of benign (MCF10A) and malignant (MCF10CA1a) breast cell lines. MCF10A-1 through MCF10A-4, and MCF10CA1a-1 through MCF10CA1a-4 represent four glycoprotein profiles from four biological replicates for each cell line. The bar at the top of the figure shows the range of expression levels (log2).

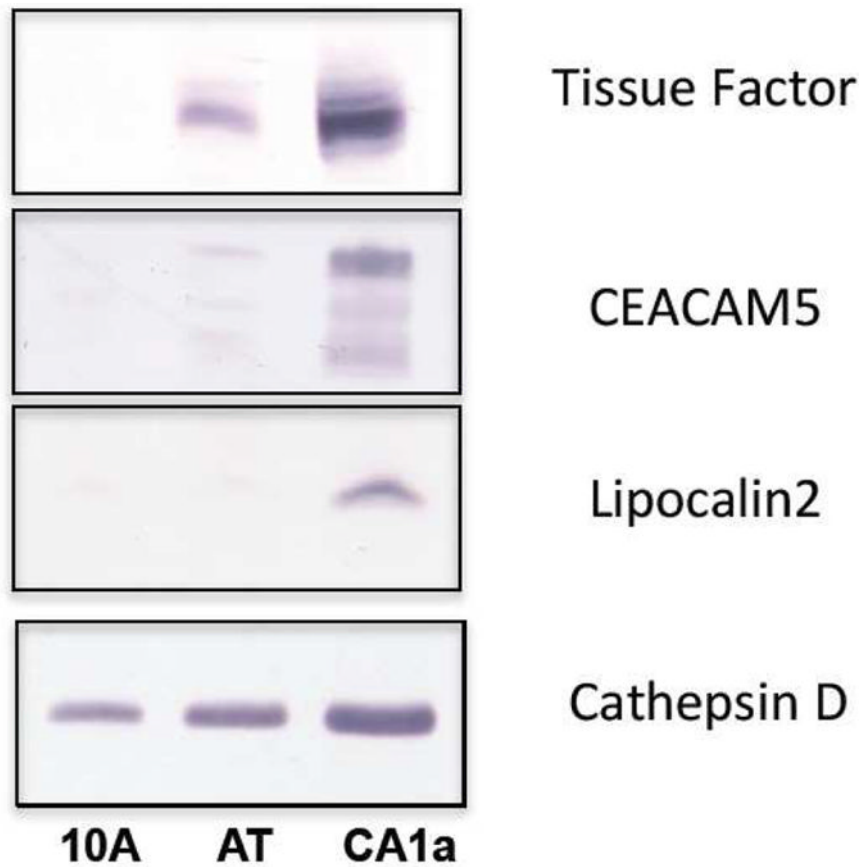


Figure 3. Expression levels of tissue factor, CEACAM5 and lipocalin2 vary among the MCF10A family of cell lines. Cathepsin D was expressed at significant levels in all three lines and is shown as a positive control for sample loading. (10A = MCF10A, AT = MCF10AT and CA1a = MCF10CA1a cell lysates).

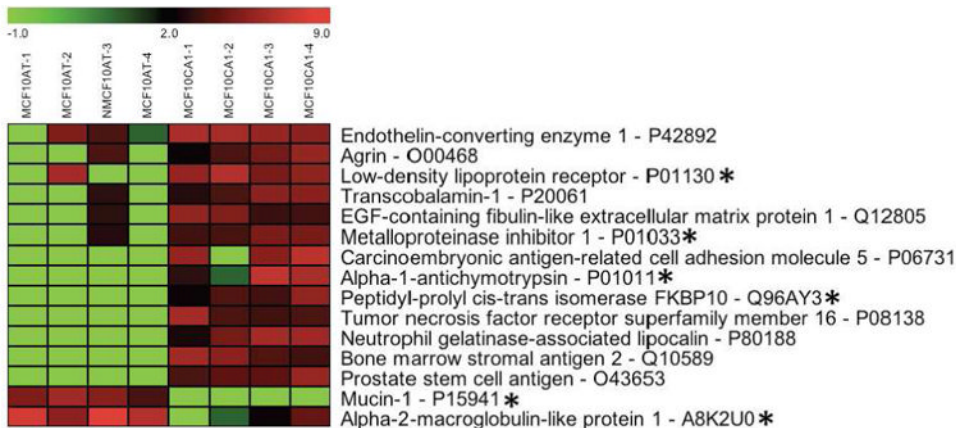


Figure 4. Differentially expressed glycoproteins comparing glycoprotein profiles of premalignant (MCF10AT) and malignant (MCF10CA1a) breast cell lines. MCF10AT-1 through MCF10AT-4, and MCF10CA1a-1 through MCF10CA1a-4 represent four glycoprotein profiles from four biological replicates for the premalignant and malignant cell lines, respectively. The gradated bar at the top of the figure shows the range of expression levels (log₂). Glycoproteins labeled with an asterisk uniquely change in the transition from premalignancy to malignancy.

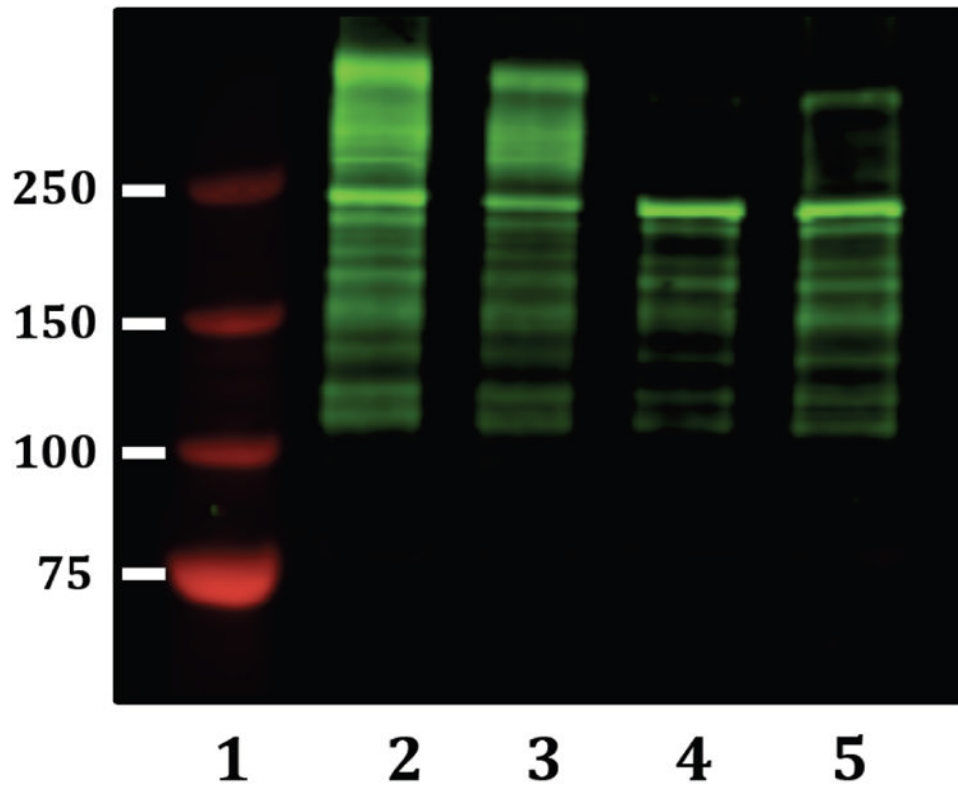


Figure 5. Western blot of normal breast and breast cancer cell lysates for collagen alpha-1(XII) chain. Lane 1, molecular weight standards (values shown on left are in kD). Lanes 2 (CP565628) and 3 (CP552484) breast tumor cell lysates. Lanes 4 (CP565401) and 5 (CP565673) normal breast tissue lysates.

Table 1

Accession Number	Gene name	Protein Name	Higher Expression in MCF10CA1		Adhesion ^c	Programmed cell death	Cell Proliferation	Cell migration	Hallmarks of Cancer				Cancer Progression
			MCF10A ^b	MCF10CA1 ^b					Angiogenesis	Invasion	Metastasis	Epithelial to Mesenchymal Transition	
P13728	F3	Tissue factor	5	35	x		x	x					x
P42892	ECE1	Endothelin-converting enzyme 1	1	11		x							
P13888	CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1	2	24		x							
P08582	MF12	Melanotransform	1	12			x	x					
P18084	ITGB5	Integrin beta-5	8	16	x				x	x			x
P08731	CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	0	0		x			x	x			x
Q9UBG0	MRC2	C-type mannose receptor 2	6	18	x				x	x			
P80188	LCN2	Neutrophil gelatinase-associated lipocalin (Lipocalin 2)	7	0	x				x	x			x
Q14672	ADAM10	Disintegrin and metalloproteinase domain-containing protein 10	6	18	x		x						x
Q10589	BST2	Bone marrow stromal antigen 2	5	0	x				x				
P08138	NGFR	Tumor necrosis factor receptor superfamily member 16	0	0			x		x				
Q43653	PSCA	Prostate stem cell antigen	3	0									
Q12805	EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	2	2		x							x
P20061	TCN1	Transcobalamin-1	0	2									
P15151	PVR	Poliovirus receptor	5	11		x							
Q00488	AGRN	Agrin	2	3					x				
QBNBJ4	G0LM1	Golgi membrane protein 1	3	8									
Q03405	PLAUR	Urokinase plasminogen activator surface receptor	5	11									x
Q04900	CD164	Sialomucin core protein 24	0	0		x							x
Q8Y337	KLK5	Kallikrein-5	0	1									
Q6UXD5	SEZOL2	Seizure 6-like protein 2	0	3									
Q88715	COL12A1	Collagen alpha-1(XII) chain	0	3									x

Higher Expression in MCF10AT and MCF10CA1^d

Accession Number	Protein Name	MCF10A				MCF10AT				MCF10CA1				Hallmarks of Cancer												
		MCF10A	MCF10A	MCF10A	MCF10A	MCF10AT	MCF10AT	MCF10AT	MCF10AT	MCF10CA1	MCF10CA1	MCF10CA1	MCF10CA1	Adhesion ^c	Programmed cell death	Cell Proliferation	Cell migration	Angiogenesis	Invasion	Metastasis	Epithelial to Mesenchymal Transition	Cancer Progression				
O95274	LYPD3		3	81	125																					
P18584	ITGB6		5	59	76																					
O43278	SPINT1		6	34	60																					
P48960	CD87		5	27	28																					
Q8N271	PROM2		0	29	21																					
Q92692	PVRL2		2	13	21																					
Q60449	LY75		0	15	17																					
O15031	PLXNB2		2	9	14																					
P00533	EGFR		0	9	13																					
Q16790	CA9		0	7	11																					
P19256	CD58		0	6	9																					
			Lower Expression in either MCF10AT and/or MCF10CA1																							
Accession Number	Protein Name	MCF10A	MCF10AT	MCF10CA1	MCF10A	MCF10AT	MCF10CA1	MCF10A	MCF10AT	MCF10CA1	MCF10A	MCF10AT	MCF10CA1	Adhesion ^c	Programmed cell death	Cell Proliferation	Cell migration	Angiogenesis	Invasion	Metastasis	Epithelial to Mesenchymal Transition	Cancer Progression				
P08571	CD14		29	6	0																					
P05362	ICAM1		21	0	0																					
Q9H3G5	CPVL		14	0	0																					
Q9UQC9	CLCA2		14	12	0																					
PGM73	HME		9	0	0																					

^a Identified by Significance analysis for microarrays (SAM) program

^b Average spectral counts from four biological replicates.

^c An X indicates that the glycoprotein has been associated with the specific hallmark of cancer.

Table 2

Glycoprotein Expression in Breast Tissue Lysates

Protein Name	Accession Number	Cancer CP565605	Cancer CP565628	Cancer CP552484	Cancer CP544778	Normal CP565673	Normal CP565401
Ly6/PLAUR domain-containing protein 3	095274	5					
Neutrophil gelatinase-associated lipocalin	P80188	3					
Bone marrow stromal antigen 2	Q10589	2	11				
C-type mannose receptor 2	Q9UBG0	2					
Aggrin	O00468	3					
Carcinoembryonic antigen-related cell adhesion molecule 5	P06731		16				
Collagen alpha-(XII) chain	Q99715	87	142	81	48	12	2
EGF-containing fibulin-like extracellular matrix protein 1	Q12805	1	6	4	3	4	1
Alpha-1-antichymotrypsin	P01011	63	140	36		96	225
Intercellular adhesion molecule 1*	P05362	92	4				
Mucin-1*	P15941		50	22			

* Expressed at higher levels in the MCF10A line (derived from benign tumor)