Systemic alteration of cell-surface and secreted glycoprotein expression in malignant breast cancer cell lines

Leslie C Timpe[1], Roger Yen, Nicole V Haste, Christina Litsakos-Cheung, Ten-Yang Yen, and Bruce A Macher

Department of Chemistry and Biochemistry, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA

Breast cancer cell lines express fewer transmembrane and secreted glycoproteins than nonmalignant ones. The objective of these experiments was to characterize the changes in the expression of several hundred glycoproteins quantitatively. Secreted and cell-surface glycoproteins were isolated using a glycoprotein capture protocol and then identified by tandem mass spectrometry. Glycoproteins expressed by a group of cell lines originating from malignant tumors of the breast were compared with those expressed by a nonmalignant set. The average number of spectral counts (proportional to relative protein abundance) and the total number of glycopeptides in the malignant samples were reduced to about two-thirds of the level in the nonmalignant samples. Most glycoproteins were expressed at a different level in the malignant samples, with nearly as many increasing as decreasing. The glycoproteins with reduced expression accounted for a larger change in spectral counts, and hence for the net loss of spectral counts in the malignant lines. Similar results were found when the glycoproteins were studied via identified glycosylation sites only, or through identified sites together with non-glycopeptides. The overall reduction is largely due to the loss of integrins, laminins and other proteins that form or interact with the basement membrane.

Keywords: breast cancer / glycoproteins / glycosylation sites / mass spectrometry / proteomics

Introduction

Biomarkers detectable in the blood would be very desirable for the purposes of diagnosing cancer, evaluating prognosis or for predicting the response to drugs, yet at present these markers are mainly used for the less demanding task of monitoring for recurrence after cancer has already been diagnosed. Glycosylated proteins that are either secreted or shed from carcinoma cells may be detectable in the blood, and are potential biomarkers. Most cancer serum biomarkers that are in clinical use are glycoproteins, including CA125 (ovarian cancer), prostate-specific antigen and carcinoembryonic antigen. The rate of approval of new biomarkers for clinical use by the US Food and Drug Administration is low, and in fact fell between 1994 and 2005 (Ludwig and Weinstein 2005). In an effort to identify a large pool of candidate biomarkers for breast cancer, Yen et al. (2012) isolated 486 N-linked cell-surface and secreted glycoproteins from a set of 14 breast cell lines. The experimental approach used a glycoprotein capture method employing hydrazide chemistry and mass spectrometry (MS) for peptide identification (Zhang et al. 2003; Wollscheid et al. 2009). The dataset included glycoproteins that are expressed differentially in cell lines of nonmalignant vs malignant origin, or of basal (myoepithelial) vs luminal origin. That dataset has been expanded subsequently to include a total of 19 cell lines. There is a quantitative difference in total glycoprotein expression found between nonmalignant and malignant lines, with malignant lines having, on average, lower levels, as described below. This observation is new. Detecting the difference in glycoprotein expression requires a systems approach in which the expression of large numbers of proteins may be measured using MS, using a fairly homogeneous source of protein such as cell lines rather than tissue samples, and having a dataset with a significant number of both nonmalignant and malignant samples.

The observation of reduced glycoprotein expression in malignant cell lines may enhance our understanding of the transition from the nonmalignant to the malignant phenotype. In the transition to a malignant state, the adherent, tight and gap junctions between the normal cells of an epithelium are weakened or lost (Martin and Jiang 2009; Paredes et al. 2012). Epithelial cells may change their morphologies and patterns of gene expression to resemble those of mesenchymal cells, thereby acquiring the ability to migrate (Thiery 2002). Extracellular proteases become active, allowing the stroma to be remodeled to accommodate the expanding tumor (Lu et al. 2011). For metastasis to occur, the carcinoma cells must be able to intravasate, extravasate and finally to form a viable colony in a tissue different from breast epithelium (Weinberg 2007). Many of these cellular processes depend on secreted proteins, or on proteins with extracellular domains, most of which are glycoproteins. Thus, an overall loss of glycoproteins could contribute to the transition from the nonmalignant to the malignant condition by

---

[1]To whom correspondence should be addressed: e-mail: lct@sfsu.edu

reducing the abundance of many proteins necessary for a cell to maintain its normal functions and its normal relations with its epithelial neighbors. The experiments to be described here make a quantitative comparison of N-linked glycoproteins between the nonmalignant and malignant breast cell lines. The comparison is made at the level of entire samples, at the level of glycoproteins and finally at the level of individual sites for N-linked glycosylation.

Results

The results described here are from a study designed to compare the cell-surface and secreted glycoprotein profiles of normal breast epithelial cells or benign tumors with those of malignant tumors, resulting in datasets for glycoproteins (Supplementary data, Table SI) and N-linked glycosylation sites (Supplementary data, Table SII). In the first step of the protocol, the glycans covalently attached to proteins of intact cultured cells were oxidized with periodate; hence, the proteins identified are either secreted or have glycosylated extracellular domains (Figure 1). Following lysis of the cells with a nonionic detergent, the oxidized glycans were coupled to hydrazide-conjugated magnetic beads for further processing. The bound proteins were digested with trypsin, and the unbound tryptic peptides (non-glycopeptides) were collected for shotgun proteomics analysis. A series of one-dimensional liquid chromatography (1D-LC) coupled with gas-phase fractionations and two-dimensional liquid chromatography electrospray ionization/tandem mass spectrometry (ESI/MS/MS) analyses were used to obtain spectra for the tryptic peptides, and the spectra

compared with databases for protein identification. The peptides covalently coupled to beads were then released by digestion with peptide N-glycosidase F (PNGase F) treatment and analyzed by a series of 1D-LC ESI/MS/MS runs. Digestion with PNGase F resulted in an increase of the mass of the N-linked glycopeptide by ~1 Da, converting asparagine to aspartate at that N-linked site and providing evidence that the asparagines in the peptides were in fact glycosylated. In this report, *glycopeptide* refers to the peptides released by PNGase F. Glycoproteins are identified both by glycopeptides and by tryptic peptides from the glycoprotein (McDonald et al. 2009). The proteins in this dataset are all annotated as glycoproteins in the UniProt database. For both the glycoprotein and glycopeptide datasets, proteins with fewer than 10 spectral counts summed over all cell lines were dropped from consideration. Since the earlier publication (Yen et al. 2012), glycoprotein datasets have been obtained for additional nonmalignant and malignant cell lines, resulting in glycoprotein and N-linked glycopeptide datasets for 19 cell types (Table I). The datasets include identifications, N-linked site locations and spectral counts for each glycopeptide and glycoprotein. Spectral counts provide a reliable means for quantitative comparison of the relative expression levels of the same glycoprotein across multiple cell samples (Old et al. 2005; Zhang et al. 2006; Zhu et al. 2010).

The initial observation was that the total number of spectral counts for glycoproteins in malignant cell lines was usually lower than in nonmalignant lines (Figure 2A). The average number of spectral counts in the malignant lines was 67% that in nonmalignant lines, with a statistically significant difference in means ($P < 5 \times 10^{-5}$, two-tailed t-test). There was also an increased spread in the distribution of spectral counts in malignant lines. There was no significant difference between nonmalignant and malignant sample protein concentrations before glycoprotein selection (1.38 vs 1.49 µg/mL, $P = 0.75$, two-tailed t-test).

Table I. Summary of cell lines
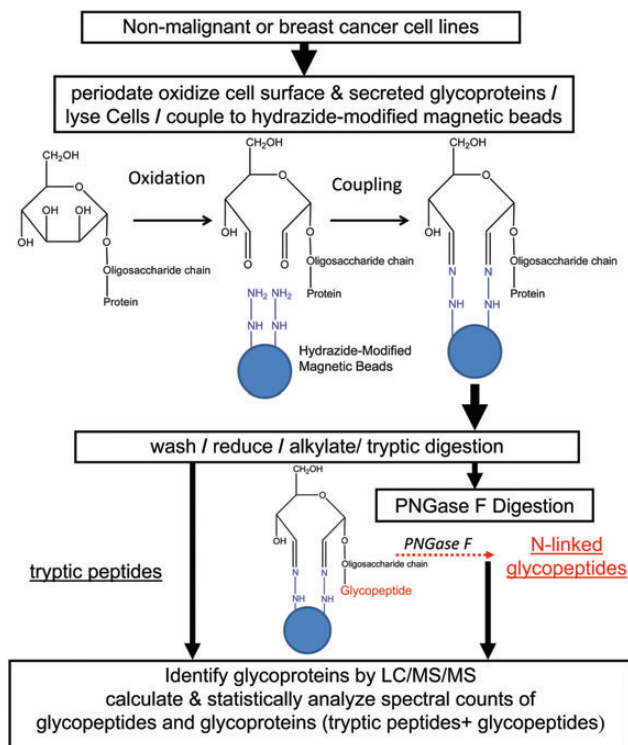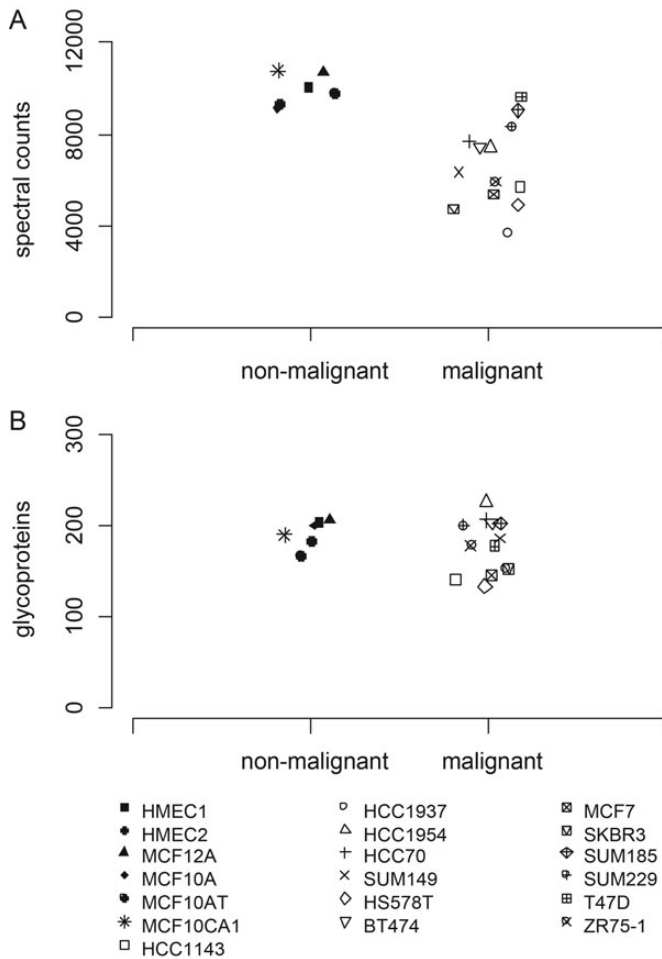
Cell line	Origin[a]	Biological replicates	Total spectral counts
HMEC 1	Human mammary epithelial cells	2	10 029
HMEC 2	Human mammary epithelial cells	3	9335
MCF12A	Fibrocystic disease	2	10 700
MCF10A	Fibrocystic disease	4	9135
MCF10AT	Ras transfected MCF10A	4	9783
MCFCA1	Subclone of MCF10AT	4	10 780
HCC1143	Ductal carcinoma	1	5708
HCC1937	Ductal carcinoma	2	3696
HCC1954	Ductal carcinoma	2	7398
HCC70	Ductal carcinoma	2	7693
SUM149	Inflammatory ductal carcinoma	2	6324
HS578T	Invasive ductal carcinoma	2	4909
BT474	Invasive ductal carcinoma	2	7436
MCF7	Invasive ductal carcinoma	2	5366
SKBR3	Adenocarcinoma	2	4726
SUM185	Adenocarcinoma	2	9059
SUM229	Pleural effusion[b]	2	8343
T47D	Invasive ductal carcinoma	2	9634
ZR751	Invasive ductal carcinoma	2	5914

[a]The information on origin is from Neve et al. (2006).
[b]Forozan et al. (1999).

Fig. 1. Outline of the protocol.

A

B

| | | | |
|---|---|---|
| ■ HMEC1 | ◌ HCC1937 | ⊠ MCF7 |
| ● HMEC2 | △ HCC1954 | ⊠ SKBR3 |
| ▲ MCF12A | + HCC70 | ⊕ SUM185 |
| ◆ MCF10A | × SUM149 | ⊕ SUM229 |
| ● MCF10AT | ◇ HS578T | ⊞ T47D |
| ✳ MCF10CA1 | ▽ BT474 | ⊠ ZR75-1 |
| ▢ HCC1143 | | |

Fig. 2. Comparison of total spectral counts or glycoproteins detected in nonmalignant vs malignant cell lines. (**A**) The spectral counts were summed over all glycoproteins detected in a cell line, and the sums are plotted. The cell-line symbols have been jittered in the horizontal direction. (**B**) The numbers of glycoproteins having at least 10 spectral counts are about the same between malignant and nonmalignant samples.

Glycoproteins
Do all glycoproteins in the malignant samples experience a reduction in spectral counts of about one-third, or does the size of the effect vary among different glycoproteins? Although the total spectral counts differed between the nonmalignant and malignant cell lines, the number of glycoproteins identified was similar (Figure 2B). All the glycoproteins in this study had a total of at least 10 spectral counts when summed over all cell lines. About the same number of glycoproteins satisfied this condition in both malignant and nonmalignant cell lines. Hence, the reduction in total spectral counts is not due to the complete loss of expression of a significant number of glycoproteins by the malignant lines.

Differences in expression at the level of individual glycoproteins were examined by calculating the difference in means, $m_m - m_{nm}$, where m_m is the average of spectral counts in the 13 malignant cell lines and m_{nm} is the average in the six nonmalignant lines. The distribution of differences for the 462 glycoproteins shows a peak close to zero, but also a very large

A

B
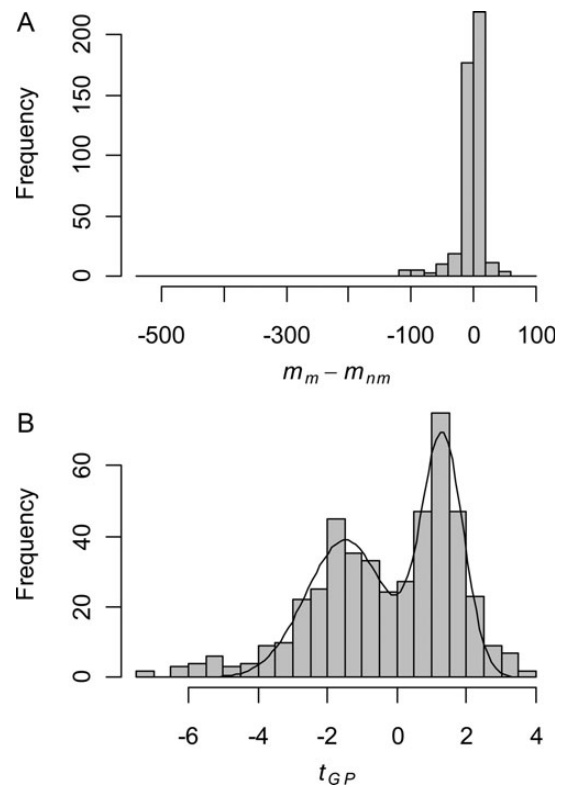
Fig. 3. Changes in glycoprotein expression between nonmalignant and malignant cell lines analyzed individually for the 462 glycoproteins in the dataset. (**A**) For each glycoprotein, the difference in mean expression between malignant and nonmalignant samples, $m_m - m_{nm}$, was calculated; the figure shows the frequency distribution of these differences. (**B**) The frequency distribution of the homoscedastic two-sample t statistics has two peaks, one corresponding to glycoproteins with higher expression, the other to glycoproteins with lower expression in the malignant samples. The curve is the sum of two normal distributions, with means -1.48, 1.31 and standard deviations 1.15, 0.60.

range (Figure 3A). The largest change, a loss of over 500 spectral counts, occurred for aminopeptidase N cluster of differentiation (CD13). A two-sample t statistic, t_{GP}, expresses the difference in spectral counts as a unitless multiple of the standard error of the difference between means. The distribution of the 462 t_{GP}'s shows two peaks, one (at $t_{GP} = 1.31$) corresponding to glycoproteins with more spectral counts in the malignant samples than in the nonmalignant ones, and the other peak ($t_{GP} = -1.48$) of glycoproteins with fewer spectral counts in the malignant lines (Figure 3B). The distribution can be fit with the sum of two normal distributions with similar areas (Figure 3B, smooth curve). At both the high and the low ends of the distribution, there are more glycoproteins than expected from the fitted curve, suggesting the presence of more than two normal components. In particular, there are several glycoproteins with $t_{GP} < -4$, which are expressed at a much lower level in malignant than in nonmalignant cell lines. Tables II and III list the top and bottom dozen glycoproteins, sorted by the difference in spectral counts between malignant and nonmalignant samples and by t_{GP}. The number of glycoproteins for which $t_{GP} < 0$, 225, is nearly the same as those for which $t_{GP} > 0$, 237. Nevertheless, the change in spectral counts is greater for the glycoproteins

Table II. The 12 glycoproteins with the largest decreases and the largest increases in spectral counts

Protein name	UniProt accession number	Difference in spectral counts[a]
Aminopeptidase N (CD13)	P15144	−533
CD59 glycoprotein	P13987	−306
Integrin beta-1 (CD29)	P05556	−299
Proactivator polypeptide	P07602	−270
Basigin (CD147)	P35613	−234
CD109 antigen	Q6YHK3	−208
CD63 antigen	P08962	−118
Lysosome-associated membrane glycoprotein 2	P13473	−109
Lysosome-associated membrane glycoprotein 1	P11279	−106
5′-nucleotidase	P21589	−105
CD44 antigen	P16070	−100
Integrin alpha-3	P26006	−97
Apolipoprotein D	P05090	28
Transmembrane emp24 domain-containing protein 10 domain-containing protein 10	P49755	31
Carbonic anhydrase 12	O43570	33
Neural cell adhesion molecule 2	O15394	33
Glucosidase 2 subunit beta	P14314	36
Alkaline phosphatase, placental type	P05187	38
Liver carboxylesterase 1	P23141	43
Galectin-3-binding protein	Q08380	50
Acid ceramidase	Q13510	52
Basal cell adhesion molecule	P50895	57
Sushi domain-containing protein 2	Q9UGT4	66
Lysosomal alpha-glucosidase	P10253	86

[a]For each glycoprotein $m_m - m_{nm}$ is provided.

Table III. The 12 glycoproteins with the largest decreases and the largest increases in t

Protein name	UniProt accession number	t_{GP}
Integrin alpha-3	P26006	−7.3
CD59 glycoprotein	P13987	−7.14
CD151 antigen	P48509	−6.78
5′-nucleotidase	P21589	−6.07
CUB domain-containing protein 1	Q9H5V8	−6.07
Poliovirus receptor-related protein 1	Q15223	−5.92
CD109 antigen	Q6YHK3	−5.85
Basigin	P35613	−5.8
Dipeptidyl peptidase 1	P53634	−5.76
Low-density lipoprotein receptor	P01130	−5.41
Integrin beta-1	P05556	−5.37
Laminin subunit alpha-3	Q16787	−5.33
UPF0577 protein KIAA1324	Q6UXG2	2.82
Lysosomal acid phosphatase	P11117	2.86
Podocalyxin	O00592	2.97
Choline transporter-like protein 2	Q8IWA5	3.06
Basal cell adhesion molecule	P50895	3.14
Tetraspanin-15	O95858	3.26
Immunoglobulin superfamily member 3	O75054	3.39
Sortilin	Q99523	3.41
Seizure 6-like protein 2	Q6UXD5	3.44
Dystroglycan	Q14118	3.48
Gamma-glutamyltranspeptidase 1	P19440	3.51
Clusterin	P10909	3.61

with reduced expression. For glycoproteins with $t_{GP} < 0$, the total difference in spectral count means is −4835, whereas for glycoproteins with $t_{GP} > 0$ the difference is 1595. To summarize, in the malignant cell lines, the spectral counts decrease in about half of the glycoproteins, while they increase in the remainder. The population of glycoproteins that has lower spectral counts in the malignant lines includes many cases in which the reductions are quantitatively large, so that there is a net loss of spectral counts in the malignant samples.

The changes in individual protein levels are also apparent in experiments using a different technique, western blot (Figure 4). Basal cell adhesion molecule is expressed at a higher level, on average, in malignant samples, as illustrated in the first row of the figure with human mammary epithelial cell (HMEC) (non-tumor cells), and SUM185 or HCC1954 (from malignant tumors). Spectral counts from MS are provided beneath the protein bands for comparison. The lower three rows are from proteins expressed at lower levels in malignant samples. Two of the glycoproteins shown, basal cell adhesion molecule and CD29, are among the glycoproteins with the largest changes in mean expression levels as listed in Table II. In general, there is good agreement between results using MS and western blot.

Glycosylation sites

N-linked glycosylation sites were identified based on the presence of the consensus sequence NXS/T, where X is any amino acid other than proline, and the asparagine residue within the consensus was converted to aspartate following PNGase

F-catalyzed release of the peptide from the hydrazide resin. Is there a decrease in the frequency of glycosylation of N-linked sites in the malignant cell lines? The dataset (Supplementary data, Table SII) contains 1037 N-linked glycosylation sites among 313 glycoproteins. The number of glycoproteins in this dataset is fewer than the 462 in the glycoprotein data because fewer glycoproteins meet the minimum threshold of 10 spectra when only N-linked glycosylation sites are counted. The average number of glycosylation sites is 3.3 per glycoprotein. The frequency distribution of the number of N-linked glycosylation sites per glycoprotein is approximately exponential, with a range of 1–27 sites (Figure 5A). Sixty-three glycoproteins had a single site. The protein with 27 glycosylation sites is pro-low-density lipoprotein receptor-related protein 1 (Q07954).

Differences between the nonmalignant and malignant samples were also observed in the glycosylation site data. The total number of glycosylation sites identified was reduced, with the malignant lines having 64% of sites identified in the nonmalignant lines (Figure 5B). The glycoproteins in this dataset met the criterion of at least 10 spectral counts in total, but some individual glycosites had very few spectral counts in the nonmalignant samples and fell to zero in the malignant group. There were also sites with the reverse pattern, but the net result was a loss of glycosites in the malignant cell lines.

A statistic t_{GS} was calculated as before to estimate the difference between nonmalignant and malignant samples at the level of individual sites. The frequency distribution shows two peaks, one corresponding to glycosylation sites that were detected more often in the malignant samples (mean = 1.07) and another for sites detected less often (mean = −1.54) (Figure 5C). As with the glycoprotein data (Figure 3B), the two components are very well fit by normal distributions, and in
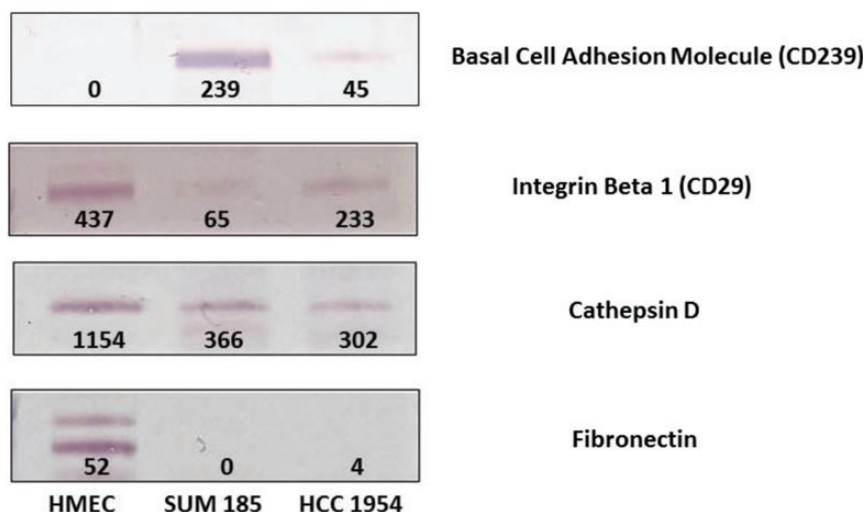
Fig. 4. Western blot analysis of four glycoproteins. The figure compares the abundances of four glycoproteins as detected by immunoblot with the corresponding spectral counts (insets). The spectral count data are from Supplementary data, Table SI.

fact have similar values for the means and standard deviations. Again the two normal components do not fit the most extreme values of t_{GS}, suggesting the presence of additional normal components in the distribution, although containing relatively few glycosylation sites.

The data collected from the breast cell lines also provide the opportunity to compare the levels of glycosylation at different N-linked sites within a glycoprotein. Scatterplots were constructed in which the points correspond to the different N-linked glycosylation sites of a given glycoprotein, with the measurements being the average of the spectral counts in the nonmalignant or malignant samples for that site. The detection of glycosylation sites within individual glycoproteins varied substantially. Glycosylation sites detected at a low frequency in the nonmalignant lines were also detected at a low frequency in the malignant lines, and similarly for high frequency sites. These scatterplots (Figure 6A and B) were usually linear, as expected if the likelihood of glycosylation occurring, or appearing in the data, is controlled by the protein structure or by the efficiency of detection by MS, rather than by altered glycosylation in malignant cells. The variability in detection probability may result from a preference for the transfer of glycans to some asparagine residues, which has been attributed to the accessibility of the asparagine within each consensus sequence (Zielinska et al. 2010; Thaysen-Andersen & Packer 2012). The slopes of the lines will change if the abundance of the glycoprotein differs between the nonmalignant and malignant samples. For example, the slope of the regression line is 2.3 for lysosomal α-glucosidase (Figure 6A). This value is in reasonable agreement with 2.9, the ratio of total counts (malignant/nonmalignant) for lysosomal α-glucosidase in the glycoprotein data. For α-3 integrin, the slope of the regression line is 0.17, compared with 0.19 in the glycoprotein data. Generally, there is a positive relationship between the ratio of malignant to nonmalignant spectral counts in the glycoprotein data and the slope of the corresponding regression lines in the glycosylation site data (Figure 6C). To summarize, the glycosite data are

consistent with the presence of two main populations of glycoproteins, one whose abundance increases and another that decreases, in the malignant cell lines compared with nonmalignant.

Transcription and gene dosage

Do changes in messenger RNA levels explain the observed differences in glycoprotein spectral counts between nonmalignant and malignant breast cell lines? Heiser et al. (2012) have analyzed mRNA expression in 55 breast cancer cell lines, 14 of which are included in the glycoprotein dataset described here, with 434 common genes/glycoproteins. The total fluorescence intensity of the 434 mRNAs does not differ between nonmalignant and malignant lines (Figure 7A), in contrast to the drop in total glycoprotein spectral counts (Figure 2A). t_{RNA} statistics for the differences between the nonmalignant and malignant lines were calculated in the same manner as for glycoproteins. The distribution of these statistics has one mode, and a tail showing an excess of extreme negative values compared with positive ones (Figure 7B). There is no evidence for two peaks, as seen for the glycoprotein data; in this respect, the protein data contain more information than the mRNA data. A scatterplot of t_{GP} vs t_{RNA} shows that there is an approximately linear association between the two variables (Figure 7C). Simple linear regression gives a slope of 0.66, which differs from 0 with a P-value of $\ll 0.01$. However, the value for r^2, 0.34, is low, meaning that the difference in gene expression between malignant and nonmalignant lines is a poor predictor of glycoprotein levels. For any change in mRNA, the change in the expression of the corresponding protein may occur over a wide range, and may be either positive or negative. If the data are viewed as categorical variables which simply count whether the RNA/glycoprotein increases or decreases, the relation between mRNA and glycoprotein changes can be summarized as in Table IV. While the glycoproteins and mRNA often vary in the same direction, they differ for 127 genes/proteins. The
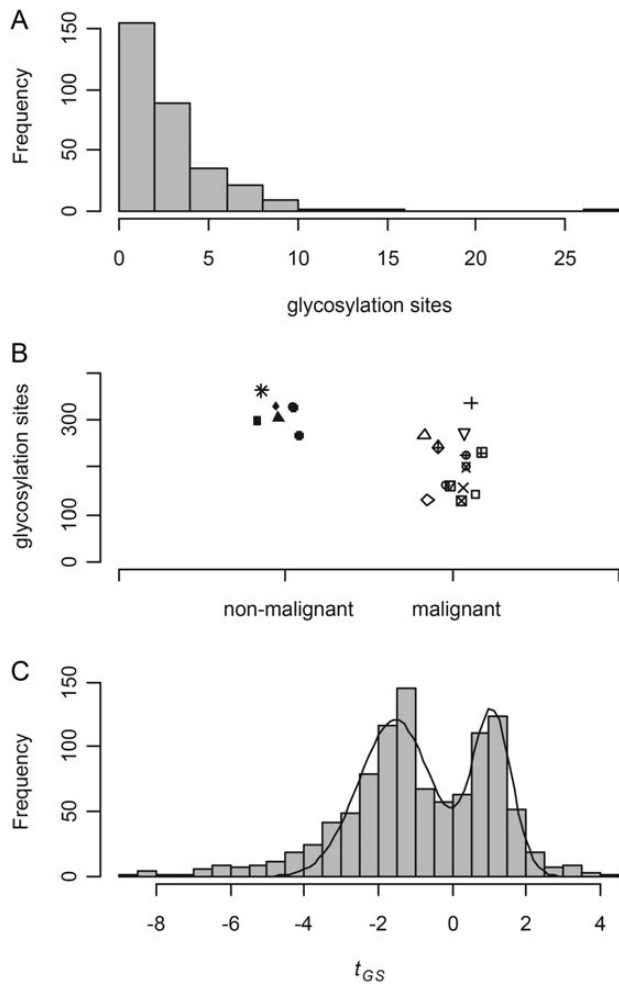
Fig. 5. Changes in the frequency of glycosylation of N-linked sites in malignant compared with nonmalignant cell lines. (**A**) Distribution of the number of N-linked sites per glycoprotein. (**B**) The number of glycosylation sites detected in the various cell lines. A glycosite was scored as detected if at least one glycopeptide contained the site. The symbols are identified by the legend in Figure 2. (**C**) Changes in spectral counts per glycosylation site calculated as t statistics. The two components correspond to sites such that the probability of being glycosylated increases (right peak) or decreases (left peak) in malignant compared with nonmalignant lines. The curve is the sum of two normal distributions, with means −1.54, 1.07 and standard deviations 0.99, 0.53.

mRNA and spectral count data are most likely to be discordant when the differences of means are small, and noise overwhelms signal in the measurements.

Heiser et al. (2012) also provide comparative genomic hybridization (CGH) data for many breast cancer cell lines. A similar comparison of t_{CGH} and t_{GP} displays no association. This result is very similar to that reported by Geiger et al. (2010). For the single case of human epidermal growth factor receptor 2, which is overexpressed in three of the cell lines and for which the gene has been copied at high multiplicity, there is a positive association between gene dosage and protein expression. However, gene dosage makes little contribution to the changes in glycoprotein expression for most of these glycoproteins.
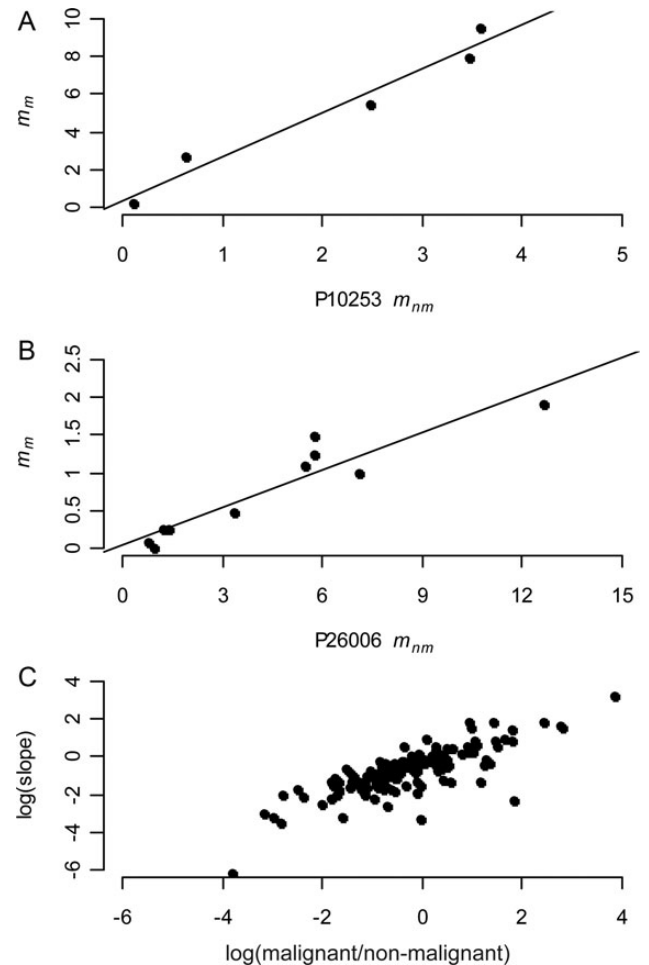
Fig. 6. Differential glycosylation site occupancy within a glycoprotein provides another method for analyzing changes in glycoprotein expression. (**A** and **B**) Scatterplots showing the association between mean spectral counts, malignant vs nonmalignant, at five glycosylation sites for lysosomal α-glucosidase (P10253) and the 10 sites for α3 integrin (P26006). (**C**) Comparison of the slopes of scatterplots as in A and B with the malignant/nonmalignant ratios from glycoprotein data. Each point corresponds to a glycoprotein with positive slope and ratio.

Discussion

These experiments demonstrate that the cell lines derived from malignant tumors have a significantly lower level of expression of cell-surface and secreted proteins than those derived from nonmalignant tissues. To our knowledge, this observation is novel. The underlying pattern of change in glycoprotein expression is more complex than a simple reduction in glycoprotein abundance, however. The distribution of t statistics shows some regularity in the data, in the form of two major peaks, suggesting a straightforward interpretation: While many glycoproteins are expressed at lower levels in the malignant samples, a large number are also expressed at higher levels. Some of the proteins with reduced expression in the malignant group have undergone a large reduction in spectral counts, accounting for the net reduction. A similar pattern is apparent in the data from glyco-
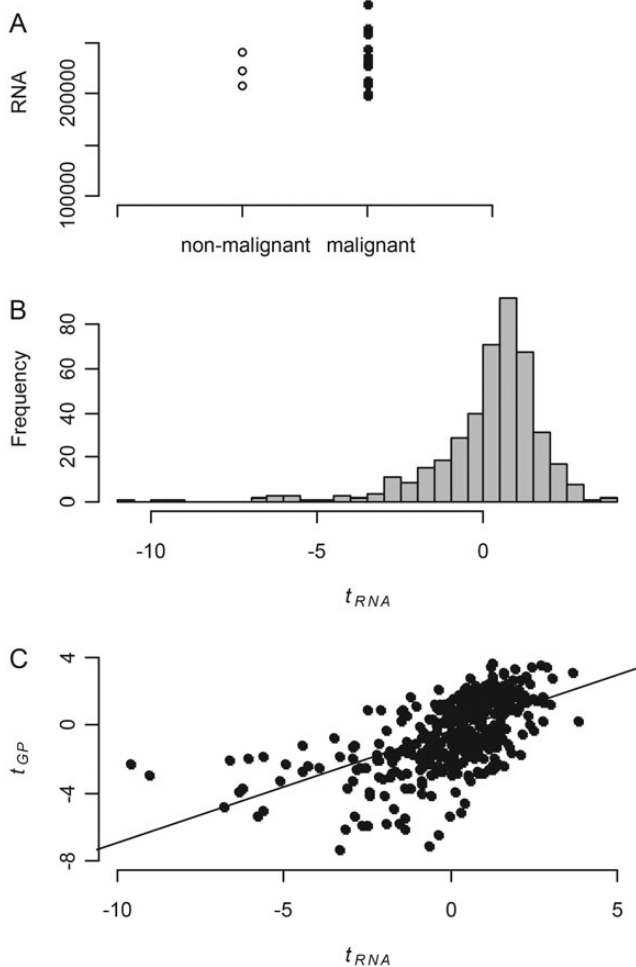
Fig. 7. Changes in glycoprotein expression as a function of changes in mRNA. The mRNA expression dataset from Heiser et al. (2012), applied to 434 glycoproteins in 14 cell lines. (**A**) The nonmalignant and malignant samples have similar total mRNA over the 434 genes. The inverse log transform has been applied to the data in Heiser et al. (2012), and the fluorescence intensities summed for all genes in each cell line. (**B**) t Statistics for mRNA show a tail to the left corresponding to reduced expression, but only a single peak. (**C**) The normalized differences for glycoproteins have a positive association with the corresponding changes in mRNA.

Table IV. Comparison of the direction of change in mRNA and glycoproteins

		Glycoprotein	
		+	−
mRNA[a]	+	190	99
	−	28	117

[a]The mRNA data are from Heiser et al. (2012). The total number of genes/proteins is 434 rather than 462 due to missing values in the RNA data.

sylation sites. The distributions of t statistics also suggest that there may be small subpopulations of glycoproteins and glycosites that undergo more extreme changes in expression.

Interpretation of changes in spectral counts
The detection of fewer spectral counts in the malignant samples raises the possibility that a reduction in glycosylation or an

Table V. Comparison of the direction of change in glycoproteins from conditioned medium and breast cancer cell lines (cell surface/ECM).

		Conditioned medium[a]	
		+	−
Cell surface	+	68	42
	−	40	90

[a]SILAC ratios from Supplemental Table 4 of Boersema et al. (2013) were transformed by the inverse log. The values from the nonmalignant cell lines were averaged for each glycoprotein, and similarly for malignant lines; these averages were compared.

alteration in the glycan structure explains these results. There are, however, several reasons to expect that the observed changes are due to alterations in the abundance of glycoproteins, rather than changes in glycosylation. First, the hypothesis of reduced glycosylation does not account for the increase in spectral counts observed for many glycoproteins. Secondly, based on the knowledge base of possible glycoprotein glycan structures in mammalian cells, the oxidation and coupling reactions will likely lead to the capture of any glycoprotein, regardless of the structural details of its glycans (Spiro 2002; Varki et al. 2009). Finally, suppose that reduced glycosylation occurs without altering protein synthesis or membrane trafficking. The resulting proteins, either lacking N-linked carbohydrates or with fewer than the normal number of sites glycosylated, would be subject to increased degradation by extracellular proteases, have shorter average lifetimes and hence be present at reduced abundances. The alternative view that glycoprotein expression can either increase or decrease in the transition to malignancy readily explains the glycoprotein data.

Other studies
There are a number of other proteomics studies of the breast cancer cell lines, but most consider only lines from malignant tumors, or cover many fewer cell lines than described here (Kulasingam and Diamandis 2007; Leth-Larsen et al. 2009; Whelan et al. 2009; Bateman et al. 2010; Drake et al. 2012; Geiger et al. 2012). An exception is a study by Boersema et al. (2013), who quantified the N-linked glycoproteins found in conditioned medium from four nonmalignant and seven malignant breast cancer cell lines. Boersema et al. (2013) used stable isotope labeling with amino acids in culture (SILAC) for relative quantitation; hence, their data cannot be compared directly with our spectral counts. However, the signs of expression-level differences can be assessed in the two studies. The overlap of Boersema et al.'s data (Supplemental Table IVB) with that presented here consists of 240 glycoproteins. Of those, 158 varied in the same direction, with 90 at lower levels in the malignant group in both datasets (Table V). This result has low probability by chance under the null hypothesis of independence ($P < 10^{-6}$, Fisher's exact test).

Both the glycoprotein and the glycosylation site data from our study show a small number of proteins or sites that have $t < -4$, corresponding to very large reductions in expression. The proteins that experience the greatest decrease in expression include several that have previously been associated with carcinogenesis. Among these are the integrins, with decreases observed for the α2, α3, α5, α6, αV, β1, β4 and β6 integrins (two others, the β3 and β5 integrins, increased). Integrins bind

various proteins of the extracellular matrix (ECM), including collagens, laminins, fibronectin, vitronectin, fibrinogen and thrombospondin. Six laminins (α3, α5, β1, β3, γ1 and γ2) were detected, all at lower expression levels in the malignant samples. Several other membrane proteins that interact with integrins or ECM components, CD44, CD63, CD111 and CD151, were also among the glycoproteins with large declines in expression in malignant cells (Tables II and III). CD44, a marker for breast cancer stem cells, binds collagen, laminin, fibronectin and hyaluronan. CD63 glycoprotein is a tetraspanin that is associated with tumor progression and that forms complexes with integrins. CD151 is another tetraspanin that is required for normal basement membrane formation. CD111 (Nectin-1, Poliovirus receptor-related protein) is a component of adherens junctions that also interacts with integrins. The reduced expression of many integrins, laminins and proteins connected to them is consistent with weaker interactions with the ECM and basement membrane, suggesting that the change in expression either reflects the transition to malignancy or contributes to causing it.

Relation to known cancer signaling pathways

The results described here demonstrate that many glycoproteins have altered expression levels in the malignant compared with the nonmalignant cells. Can these changes be understood as examples of known carcinoma cell biology or signaling pathways? Some individual changes in protein expression are in the direction expected for cells undergoing the transition from an epithelial to a mesenchymal-like state. For example, E-cadherin levels are lower in the malignant samples, whereas N-cadherin levels are higher. Evdokimova et al. (2009) showed that overexpression of Y-box binding protein-1 (YB-1) in the breast cancer cell line MCF10AT induces an epithelial-mesenchymal transition (EMT). Their Table S1 provides the fold change in mRNA level observed in cells overexpressing YB-1; 47 of these genes correspond to glycoproteins present in the current dataset. Comparing the direction of mRNA or protein change in the two datasets, however, did not show a distribution that differed significantly from the one expected by chance. Hence, the changes in protein expression are not the ones expected for an EMT, or at least as caused by YB-1 overexpression.

Estrogens and retinoic acid both regulate gene expression in breast epithelial cells. Hua et al. (2009) compared the genomic targets of retinoic acid receptors and of α-estrogen receptors in a breast cancer cell line, MCF7, and also examined the effects of retinoids on gene expression in those cells. Comparing the mRNA and glycoprotein results for the 86 shared mRNA/proteins, mRNA and glycoproteins changed in the same direction significantly more often than expected by chance (Table VI) ($P < 0.01$, Fisher's exact test). Many of the genes for these 86 glycoproteins have receptors for retinoic acid and/or estrogens in their neighborhoods (Supplemental Table 6 in Hua et al. 2009). Hence, it is possible that at least some of the observed differences in glycoprotein expression between nonmalignant and malignant cells are due to signaling by the retinoid pathway. If so, the direction of this effect is unexpected, given that retinoic acid has antiproliferative effects on tumor cells. The actions of retinoic acid and possibly estrogens on glycoprotein expression, and the consequent changes in epithelial cell function, remain to be explored.

Table VI. Comparison of the direction of change in mRNA (Hua et al. 2009) and glycoproteins in breast cancer cell lines

		Glycoprotein	
		+	−
mRNA[a]	+	20	24
	−	15	37

[a]The mRNA data are from a study by Hua et al. (2009) (Supplemental Table 3) that examined the effect of retinoic acid agonists on gene expression in MCF7 cells.

In conclusion, there are numerous reports of individual glycoproteins changing expression levels in malignant compared with nonmalignant cells or tissues. According to the results described here, in breast cancer cell lines these changes can be understood as examples of a systemic pattern of altered glycoprotein expression, in which many of the proteins decline in abundance, but many also increase. The pattern is simple, dominated by two main sets of responses, as seen in the distributions of t statistics. Those glycoproteins that decline the most include laminins, integrins and other glycoproteins that interact with the basement membrane or ECM. The reduced expression of these proteins may contribute to malignancy. These proteins with very low t values appear to be part of a component of the distribution of t values that is distinct from the two larger ones, raising the possibility that their expression is controlled differently than the majority of other glycoproteins. The reduction of cell-surface glycoproteins in the malignant samples, with no detectable reduction in total protein, indicates that perhaps the functioning of the endoplasmic reticulum and Golgi apparatus is altered in the malignant samples. If so, the effect is not a simple loss of function, as roughly half the glycoproteins display an increase rather than a decrease in expression.

Materials and methods
Mass spectrometry
The spectral count data were obtained as described in detail by Yen et al. (2012), but have been extended with data from a second human mammary epithelial cell (HMEC) culture and the established breast cancer cell lines MCF10AT, MCF10CA1, Hs578 T and HCC70 (Table I; data in Supplemental data, Tables SI and SII). The set of six nonmalignant lines comprises two HMEC cultures (from different reduction mammoplasty patients, Cell Applications, Inc.; mammary epithelial cell growth medium), MCF12A, MCF10A, MCF10AT (1:1 DMEM, Ham's F-12 supplemented with 5% New Zealand horse serum, 10 μg/mL insulin, 20 ng/mL epithelial growth factor, 0.5 μg/mL hydroxycortisone, 0.1 mg/mL cholera toxin) and MCF10CA1 (1:1 DMEM, Ham's F-12, 5% New Zealand horse serum, 100 U Penstrep). The MCF 10AT and MCF10CA1 cell lines were provided by Dr. Jose Lopez, Department of Surgery, University of California San Francisco. The MCF10A and MCF12A cell lines came originally from fibrocystic (nonmalignant) growths. The MCF10CA1 line is malignant as a xenograft in immunosuppressed mice; with respect to total spectral counts, it is similar to the other nonmalignant lines that it was derived from and is grouped with them here. The remaining 13 lines originated in tumors that were malignant in human patients. Cell lines from malignant tumors that have been added to the data

of Yen et al. (2012) are Hs578 T (from ATCC, cultured in DMEM, 10% fetal bovine serum, 10 μg/mL insulin) and HCC70 (ATCC; RPMI 1640 plus 10% fetal bovine serum). The cell-line data were gathered in one to four biological replications. Protein concentrations in cell lysates were determined by Bradford assay.

Western blot analysis

Five or 10 μg of cell lysate was loaded per lane on a 4–12% NuPAGE Bis–Tris gel (Invitrogen, Carlsbad, CA) under reducing (30 mM beta-mercaptoethanol) or nonreducing conditions. A voltage of 200 V 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% bovine serum albumin in phosphate buffered saline and then incubated with one of the following primary antibodies: Rabbit anti-basal cell adhesion molecule, CD239 (1:500 dilution of Abcam ab134110, Cambridge, MA), chicken anti-integrin beta 1, CD29 (1:2000 dilution of ProSci, Inc. antibody XW-8078), goat antifibronectin (1:1000 dilution of Biosciences, San Jose, CA, antibody 610077) or mouse anti-cathepsin D (1:1000 Dilution of ProSci, Inc., Poway, CA, antibody 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) or anti-goat-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.

Data analysis

The data analyzed are the number of identified peptide MS/MS spectra, or spectral counts. These are quantitative data that allow the relative abundances of a given glycoprotein or glycopeptide to be measured in samples from different cell lines (Old et al. 2005; Zhang et al. 2006; Zhu et al. 2010). The 462 glycoproteins in Supplementary data, Table SI, have at least 10 spectral counts from both glycopeptides and non-glycopeptides, summed over all 44 biological replicates. For the glycopeptide data (Supplementary data, Table SII), there are at least 10 spectral counts summed over all glycosylation sites of a given glycoprotein, giving 313 glycoproteins. We have examined over 800 of the spectra and estimate that the chance of an error in assignment is <15%. With a threshold of at least 10 assigned spectra, the probability of false identification of a glycoprotein is close to zero.

The t statistic used was the standard one for the difference between two sample means, unequal sample sizes, assuming the same population variance:

$$t = \frac{(m_\mathrm{m} - m_\mathrm{nm})}{s_\mathrm{m,nm}\sqrt{(1/n_\mathrm{m}) + (1/n_\mathrm{nm})}},$$

with

$$s_\mathrm{m,nm} = \sqrt{\frac{(n_\mathrm{m} - 1)s_\mathrm{m}^2 + (n_\mathrm{nm} - 1)s_\mathrm{nm}^2}{n_\mathrm{m} + n_\mathrm{nm} - 2}},$$

where s_m, s_nm, n_m and n_nm are the sample standard deviations

and sizes for the malignant and nonmalignant cell lines. These statistics were calculated over all biological replicates, rather than over the averaged data (Supplemental data, Table SI), so that the replicates have equal weight.

Supplementary data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Funding

Conflict of interest

None declared.

Abbreviations

CD, cluster of differentiation; 1D-LC, one-dimensional liquid chromatography; ECM, extracellular matrix; EMT, epithelial-mesenchymal transition; ESI, electrospray ionization; HMEC, human mammary epithelial cell; m/z, mass/charge; MS, mass spectrometry; PNGase F, peptide N glycosidase F; SILAC, stable isotope labeling with amino acids in culture; YB-1, Y-box binding protein-1'.

References

Bateman NW, Sun M, Hood BL, Flint MS, Conrads TP. 2010. Defining central themes in breast cancer biology by differential proteomics: Conserved regulation of cell spreading and focal adhesion kinase. *J Proteome Res*. 9 (10):5311–5324.

Boersema PJ, Geiger T, Wisniewski Mann M. 2013. Quantification of the *N*-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics*. 12:158–171.

Drake PM, Schilling B, Niles RK, Prakobphol A, Li B, Jung K, Cho W, Braten M, Inerowicz HD, Williams K, et al. 2012. Lectin chromatography/mass spectrometry discovery workflow identifies putative biomarkers of aggressive breast cancers. *J Proteome Res*. 11:2508–2520.

Evdokimova V, Tognon C, Ng T, Ruzanov P, Melnyk N, Fink D, Sorokin A, Ovchinnikov LP, Davicioni TJ, Sorensen PHB. 2009. Translational activation of Snail and other developmentally regulated transcription factors by YB-1 promotes an epithelial-mesenchymal transition. *Cancer Cell*. 15:402–415.

Forozan F, Veldman R, Ammerman CA, Parsa NZ, Kallioneimi A, Kallioneimi O-P, Ethier SP. 1999. Molecular cytogenetic analysis of 11 new breast cancer cell lines. *Br J Cancer*. 81(8):1328–1334.

Geiger T, Cox J, Mann M. 2010. Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet*. 6(9):e1001090.

Geiger T, Madden SF, Gallagher WM, Cox J, Mann M. 2012. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Res*. 72:2428–2439.

Heiser LM, Sadanandam A, Kuo W-L, Benz SC, Goldstein TC, Ng S, Gibb WJ, Wang NJ, Ziyada S, Tong F, et al. 2012. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci USA*. 2012:2724–2729.

Hua S, Kittler R, White KP. 2009. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell*. 137:1259–1271.

Kulasingam V, Diamandis EP. 2007. Proteomics analysis of conditioned media from three breast cancer cell lines: A mine for biomarkers and therapeutic targets. *Mol Cell Proteomics*. 6(11):1997–2011.

Leth-Larsen R, Lund R, Hansen HV, Laenkholm A-V, Tarin D, Jensen ON, Ditzel HJ. 2009. Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Mol Cell Proteomics*. 8:1436–1449.

Lu P, Takai K, Weaver VM, Werb Z. 2011. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harbor Perspect Biol*. 3(12). doi:pii: a005058. 10.1101/cshperspect.a005058

Ludwig JA, Weinstein JN. 2005. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer*. 5:845–856.

Martin TA, Jiang WG. 2009. Loss of tight junction barrier function and its role in cancer metastasis. *Biochim Biophys Acta*. 1788(4):872–891.

McDonald CA, Yang JY, Marathe V, Yen T-Y, Macher BA. 2009. Combining results from lectin affinity chromatography and glycoproteins capture approaches substantially improves the coverage of the glycoproteome. *Mol Cell Proteomics*. 8:287–301.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, et al. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 10(6):515–527.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*. 4(10):1487–1502.

Paredes J, Figueiredo J, Albergaria A, Oliveira P, Carvalho J, Ribeiro AS, Caldeira J, Costa AM, Simões-Correia J, Oliveira MJ, et al. 2012. Epithelial E- and P-cadherins: Role and clinical significance in cancer. *Biochim Biophys Acta*. 1826(2):297–311.

Spiro RG. 2002. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 12(4):43R–56R.

Thaysen-Andersen M, Packer N. 2012. Site-specific glycoproteomics confirms that protein structure dictates formation of *N*-glycan type, core fucosylation and branching. *Glycobiology*. 22(11):1440–1452.

Thiery JP. 2002. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer*. 2:442–454.

Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. 2009. *Essentials of Glycobiology*, 2nd ed., www.ncbi.nlm.nih.gov/books/NBK1908/. (14 August 2013, date last accessed).

Weinberg RA. 2007. *The Biology of Cancer*. New York, NY: Garland Science.

Whelan SA, Lu M, He J, Yan W, Saxton RE, Faull KF, Whitelegge JP, Chang HR. 2009. Mass spectrometry (LC-MS/MS) site-mapping of N-glycosylated membrane proteins for breast cancer biomarkers. *J Proteome Res*. 8 (8):4151–4160.

Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, Schiess R, Aebersold R, Watts JD. 2009. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol*. 27 (4):378–386.

Yen T-V, Macher BA, McDonald CA, Alleyne-Chin C, Timpe LC. 2012. 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*. 11:656–667.

Zhang H, Li XJ, Martin DB, Aebersold R. 2003. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 21:660–666.

Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, Samatova NF. 2006. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res*. 5:2909–2918.

Zhu W, Smith JW, Huang C-M. 2010. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotech*. 2010:6 pages, Article ID 840518, doi:10.1155/2010/840518.

Zielinska DF, Gnad F, Wisniewski JR, Mann M. 2010. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 141:897–907.