analytical chemistry

ARTICLE

pubs.acs.org/ac

A Proteomics Platform Combining Depletion, Multi-lectin Affinity Chromatography (M-LAC), and Isoelectric Focusing to Study the Breast Cancer Proteome

Zhi Zeng,[†] Marina Hincapie,[*,†] Sharon J. Pitteri,[‡] Samir Hanash,[‡] Joost Schalkwijk,[§] Jason M. Hogan,[‡] Hong Wang,[‡,⊥] and William S. Hancock[†,∥]

[†]Department of Chemistry and Chemical Biology, Barnett Institute, Northeastern University, Boston, Massachusetts, United States

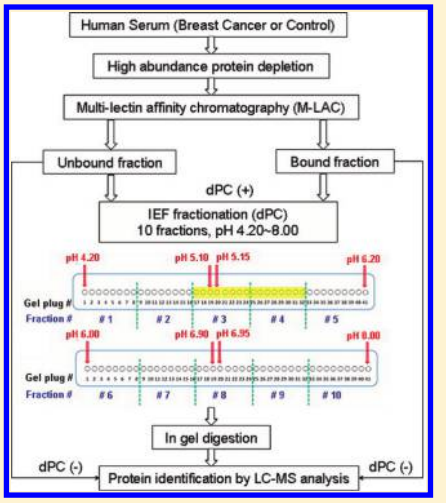
[‡]Fred Hutchinson Cancer Research Center, Seattle, Washington, United States

[§]Department of Dermatology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, The Netherlands

[∥]Department of Biomedical Science, Yonsei University, Seoul, South Korea

Supporting Information

ABSTRACT: The discovery of breast cancer associated plasma/serum biomarkers is important for early diagnosis, disease mechanism elucidation, and determination of treatment strategy for the disease. In this study of serum samples, a multidimensional fractionation platform combined with mass spectrometric analysis were used to achieve the identification of medium to lower abundance proteins, as well as to simultaneously detect glycan and abundance changes. Immuno-affinity depletion and multi-lectin chromatography (M-LAC) were integrated into an automated HPLC platform to remove high abundance protein and fractionate glycoproteins. The collected glycoproteomes were then subjected to isoelectric focusing (IEF) separation by a digital ProteomeChip (dPC), followed by in-gel digestion and LC−MS analysis using an Orbitrap mass spectrometer. As a result, the total number of identified proteins increased significantly when the IEF fractionation step was included as part of the platform. Relevant proteins with biological and disease significance were observed and the dynamic range of the serum proteome measurement was extended. In addition, potential glycan changes were indicated by comparing proteins in control and cancer samples in terms of their affinity to the multi-lectin column (M-LAC) and the p*I* profiles in IEF separation. In conclusion, a proteomics platform including high abundance protein depletion, lectin affinity fractionation, IEF separation, and LC−MS analysis has been applied to discover breast cancer-associated proteins. The following candidates, thrombospondin-1 and 5, alpha-1B-glycoprotein, serum amyloid P-component, and tenascin-X, were selected as promising examples of the use of this platform. They show potential abundance and glycan changes and will be further investigated in future studies.



INTRODUCTION

Breast cancer is the second most frequently diagnosed cancer among women in the United States. It is estimated that 207 090 new cases of invasive and 54 010 new cases of in situ breast cancers occurred in females in the US in 2010.[1] Early detection is crucial to reduce the death rate of breast cancer, yet the symptoms are not obvious in many cases. Body fluids, such as plasma or serum are good sources of biomarkers, and therefore, efforts to develop blood-based biomarker tests for early detection of breast cancer continue to be actively pursued.

The mining of low abundance proteins as biomarkers remains challenging due to the large dynamic range of the blood proteome ($\sim 10^{10}-10^{12}$ or greater[2]). Because the complexity of the unfractionated proteome exceeds the capability of any single proteomics method, multidimensional separations could, if sample loss is limited, greatly enhance the proteome coverage.

Glycosylation is one of the most common protein post-translational modifications, and the characterization of glycan structures is expected to broaden the scope of discovery studies beyond the protein level, and thus improves the clinical values of existing biomarkers. Due to the importance of glycosylation from cellular proliferation to angiogenesis and metastasis,[3] the development and application of techniques and methodologies for enriching or fractionating the glycoproteome has become an emerging field. Lectins have been widely used in various

Received: February 27, 2011
Accepted: April 22, 2011
Published: April 22, 2011

glycoproteomic studies, such as chromatography,[4] array[5,6] and blotting studies.[7] We have previously demonstrated the use of a multi-lectin affinity chromatography (M-LAC) technique to comprehensively capture glycoproteins from biological fluids.[8,9] A recently reported improved format, namely high-performance multi-lectin affinity chromatography (HP-MLAC),[10] has been applied in this study.

Isoelectric focusing (IEF) separates proteins based on their isoelectric points and has advantages including high resolution and modest cost.[11] Separation based on p*I* differences is valuable for the study of protein post translation modifications (PTM) such as glycosylation and phosphorylation.[12] The resolving power of IEF in a proteomics study has been demonstrated by many research groups. For example, Dr. Speicher's laboratory used a four-dimensional strategy including solution IEF (MicroSol-IEF) starting with 2.4 μL serum, resulting in the identification of 575 and 2890 proteins within 140 fractions from plasma and serum, respectively.[12] Nowadays, there are several commercially available platforms, such as the Agilent OFFGEL fractionator, the Bio-Rad Rotofor system, the Invitrogen ZOOM IEF fractionator and the Cell Biosciences Digital ProteomeChip (dPC).[13] We analyzed the depleted and multi-lectin fractionated serum (glyco)proteome derived from healthy and breast cancer patients by the solution phase IEF fractionation system using the dPC. The dPC is a recently introduced technique that uses parallel IEF[14] and gel plugs for rapid (30 or 45 min) and reproducible separation of complex protein or peptide mixtures. It is also an IEF technique that is compatible with the downstream in-gel digestion and LC−MS analysis. We recently reported on the use of the dPC system to detect glycosylation changes by comparing the pI profiles of high-medium abundance glycoproteins and the integration with other analytical methods including 1D SDS-PAGE, LC−MS proteomic analysis and antibody-lectin arrays.[7] In this report, we describe the application of the dPC to fractionate breast cancer and control serum samples, resulting in the identifications of a number of low abundance or disease related (glyco)proteins from 10 IEF fractions starting with 50 μL serum.

EXPERIMENTAL SECTION

Serum Samples. Five serum samples from patients with breast cancer stage 2 and another five from healthy individuals were provided by Dr. Samir Hanash's laboratory at the Fred Hutchinson Cancer Center; age and race were matched between healthy and patient individuals; all samples have been collected with IRB approval. Individual serum was pooled to give one control and one disease samples to minimize individual variability. Five lung cancer serum samples with matched age and gender were purchased from ProteoGenex. All the samples were stored at −75 °C and did not undergo more than two freeze/thaw cycles. For the purpose of method development, 10 normal female sera were purchased from Bioreclamation (Jericho, NY) and pooled as one reference serum sample, which was subsequently aliquoted and stored at −75 °C.

High Abundance Protein Depletion and Glycoprotein Fractionation. The pooled serum was fractionated using the automated HPLC sample preparation platform, previously developed in our laboratory,[10] with some minor modifications. In this study, abundant protein removal (albumin and immunoglobulins G, M, A, E, D and free light chains) was accomplished using CaptureSelect depletion resins (BAC. B.V, Naarden, The Netherlands). The anti-albumin and anti-immunoglobulin beads were gravity packed into an omnifit glass column (10 × 20 mm and 6.6 × 20 mm, respectively). The flow through from the CaptureSelect depletion column was fractionated online using a high performance multilectin affinity chromatography (M-LAC) into an unbound (unretained proteins) and a bound (retained proteins) fractions. The lectin column consists of mixtures of lectins concanavalin A (ConA), jacalin (JAC) and wheat germ agglutinin (WGA). The details of this fractionation are described in the Supporting Information. The unbound and bound M-LAC fractions were collected and further fractionated using Isoelectric Focusing (IEF).

Isoelectric Focusing (IEF) Analysis using Digital ProteomeChips (dPC). The total amount of protein loaded on the dPC chip was first optimized. For this purpose, different amount of M-LAC bound proteins (10, 20, 30, 50, 80, and 110 μg) derived from healthy human serum were applied to a dPC (Cell Biosciences, Inc., Santa Clara, CA). Sample preparation and separation in IEF are described in the Supporting Information.

LC-MS. Clinical samples were analyzed by LTQ-Orbitrap linear ion trap hybrid mass spectrometer (ThermoFisher Scientific, Waltham, MA). The nano-LC separation was performed using an Eksigent 1D+ nano-LC system (Dublin, CA). The LC method for the separation of digested peptides is described in the Supporting Information. In mass spectrometry analysis, the ion isolation width was 2 m/z; Orbitrap FullMS target value was 1×10^6; Ion trap MSn target value was 1×10^4. Xcalibur MS acquisition software (Xcalibur 2.0.7, ThermoFisher Scientific) was used for MS/MS analysis in data dependent mode. The scan cycle includes a full scan MS acquisition in profile mode at the resolution of 60K over the range of 400−1800 m/z, followed by up to 5 data-dependent MS/MS scans. Parameters in dynamic exclusion were set as follows: repeat count, 1; repeat duration, 30 s.; exclusion list size, 100; exclusion list duration, 45 s; exclusion mass width, 0.55 m/z low and 1.55 m/z high. The activation of parallel operation of LTQ and Orbitrap was achieved by the FTMS preview scan. Only +2 and +3 peptides were interrogated by MS/MS.

Bioinformatics and the Selection for Protein Candidates. MS/MS spectra were searched against the human database using the Mascot software (Matrix Sciences).[15] The parameters for search were set as follows: carbamidomethylation (C) as fixed modification; maximum missed cleavage sites was 1; mass values were using monoisotopic peaks; peptide and fragment mass tolerances were ±5 ppm and ±0.8 Da, respectively; monoisotopic peaks with one ^{13}C were taken into account. Peptides with p values smaller than 0.005 and scores higher than 25 were considered. Decoy database searching was also included to evaluate the false positive rate (FPR). The resulting peptide list of each fraction was uploaded onto Mass Spec Results Analysis Tool (MSRAT) Bioinformatics Software (Cell Biosciences, Inc., Santa Clara, CA) for the comparison of the p*I* profiles between control and disease samples. Criteria used to select protein candidates can be found in the Supporting Information.

ELISA of Tenascin-X (TN-X). Serum TN-X was measured by a sandwich type ELISA. Five control and five breast cancer serum samples were used in this ELISA study. They are the same samples that were pooled together and used in the discovery stage. Affinity purified rabbit antibodies against recombinant TN-X FNIII repeats 27 to 32 were used for antigen capture. These antibodies recognize several C-terminal fragments and full length TN-X that present in human serum.[16] The secondary antibody of the ELISA was raised in guinea pig, against recombinant protein from the C-terminal 100 kDa portion of

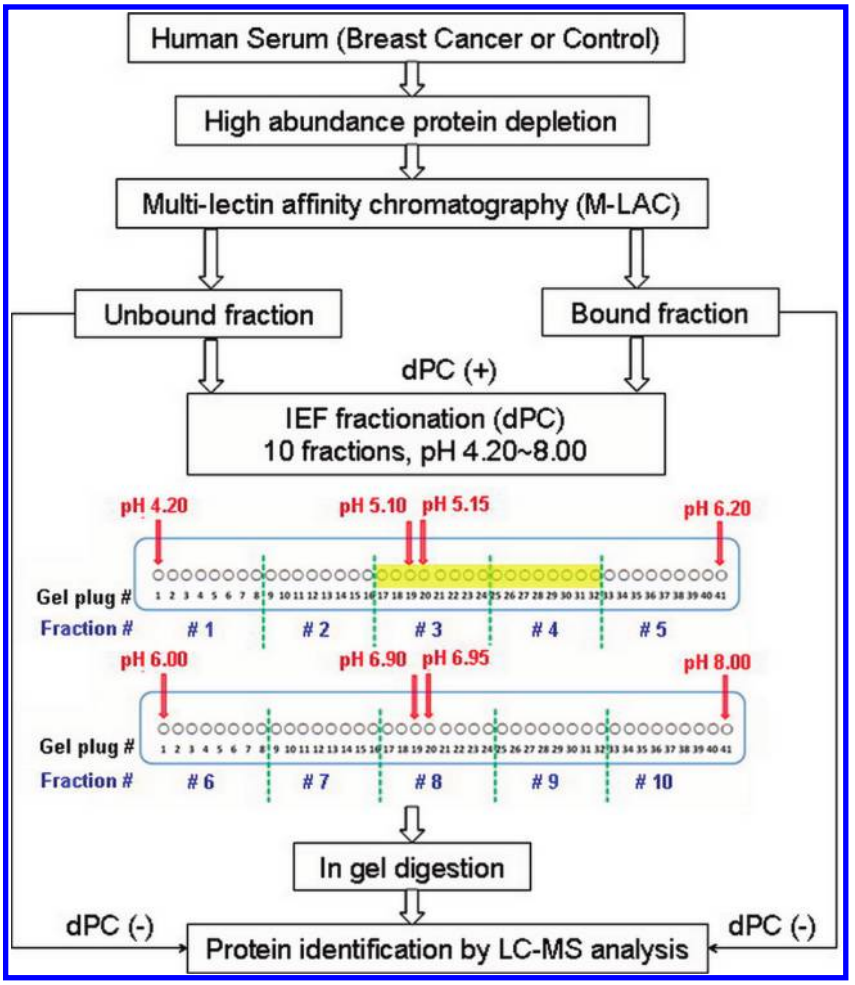


Figure 1. Diagram of the workflow used for the analysis of the breast cancer and control sera. The strategy consists of four separation approaches—depletion, glycoprotein fractionation, IEF fractionation using dPC and RP-LC−MS peptide separation. Gel plugs in yellow were chosen for dPC capacity study (Figure S-1, Supporting Information). Analysis platform with or without IEF separation is defined as dPC(+) or dPC(−) approach, respectively.

TN-X, which comprises FNIII repeats 27−32 and the fibrinogen-like domain. As there are fragments of many different sizes (presumable splice forms or breakdown products) present in human serum, no meaningful figure can be given for the total TN-X concentration in serum. For this reason, a calibration curve of recombinant TN-X FNIII repeats 27 to 32 was used. The TN-X serum levels given here are equivalent to ng/mL of this recombinant TN-X preparation. The ELISA is highly specific, as the amount of TN-X detected is zero in serum derived from patients that are homozygous for TN-X null alleles.[17] In addition, this ELISA does not detect tenascin-C, another member of the tenascin gene family.

RESULTS AND DISCUSSION

Description of the Proteomic Analysis Workflow. As indicated in the workflow in Figure 1, high abundance protein depletion was used as the first dimension; lectin fractionation and IEF were applied as the second and the third dimensions, respectively. As indicated in the workflow in Figure 1. The contribution of the depletion and lectin fractionations to increase the number of protein identification and overall sensitivity has been demonstrated in previous publications.[18,19] Approximately 80% of the high abundance protein representing albumin and immunoglobulins were initially removed by two immunodepletion columns. To capture the serum glycoproteins comprehensively, a mixture of lectins (ConA, WGA and JAC) with affinity for a broad range of different glycan structures (mannose, glucose, *N*-acetylgalactosamine in O-linked glycan, sialic acid and *N*-acetylglucosamine) was used. The specificity and reproducibility of this multi-lectin column has been demonstrated in a previous publication.[10] The glycoproteins with strong specificity toward the M-LAC column were retained as the bound fraction and later eluted at low pH; nonglycosylated proteins and glycoproteins with low levels of binding to the multi-lectin column were collected as the unbound fraction. After the removal of albumin and immunoglobulins, the remaining serum proteins distributed in M-LAC unbound and bound fractions at a ratio of about 1:1 (w/w).

As shown in Figure 1, the M-LAC unbound and bound proteomes were subsequently subjected to further fractionation using the digital ProteomeChip (dPC). Chips with operating ranges of pH 4.20−6.20 and pH 6.00−8.00 were used to separate proteins into 82 gel plugs with a pH 0.05 variation between adjacent plugs. We then combined the gel plugs into ten fractions (Figure 1) for in-gel digestion and LC−MS analysis. As a

dx.doi.org/10.1021/ac2002802 |*Anal. Chem.* 2011, 83, 4845–4854

Table 1. Comparison of dPC(+) and dPC(−) Approaches in Terms of Total Protein Hits

Protein description	dPC(+)[a]	dPC(−)[a]
Alpha-2-antiplasmin	376	22
Angiotensinogen	472	14
Biotinidase	64	2
Ceruloplasmin	2927	181
Fibronectin	2657	100
Heparin cofactor 2	492	11
Histidine-rich glycoprotein	542	68
Insulin-like growth factor-binding protein complex acid labile chain	257	9
Interalpha-trypsin inhibitor heavy chain H2	1158	94
Interalpha-trypsin inhibitor heavy chain H3	206	6
Kallistatin	192	8
Lipopolysaccharide-binding protein	110	3
Pregnancy zone protein	2186	208
Serum paraoxonase/arylesterase 1	281	10
Zinc-alpha-2-glycoprotein	284	11

[a] The sum of spectral counts in duplicate runs.

negative control, M-LAC fractions without dPC separation were also subjected to trypsin digestion and LC−MS analysis, which was defined as the dPC(−) approach. The dPC(+) platform is able to "mine" the (glyco)proteome with the advantage of additional information about potential glycan alterations as described below. In addition, the loading capacity has been optimized to be 50 μg (Figure S-1, Supporting Information) and the reproducibility of the dPC has been demonstrated (Figure S-2, Supporting Information). As for the throughput of the platform, the fractionation for a single sample was accomplished in approximately 80 min (40 min for depletion and multilectin chromatography; 30 or 45 min for IEF separation). For convenience, trypsin digestion was performed overnight. The LC−MS analysis of the resulting peptides from one IEF fraction required 4 h including peptide separation, blank and equilibration. Thus, compared to other platforms commonly used in the discovery stage, our approach compares well in terms of time required for sample depletion, glycan fractionation and IEF or ion exchange separation.

Further Fractionation using dPC for Higher Confidence in Protein Identification. The improved performance of our serum proteomics platform with incorporation of the dPC system, dPC(+), is shown in Table 1 by the identification of a selection of 15 abundant serum proteins with an increased detection signal (spectral count) in comparison with the dPC-(−) system. Also, the observed proteins were detected with increased sequence coverage (data not shown), which resulted in a higher confidence of protein identification. As can be seen from the table, about half of the proteins were observed with a 10-fold or more increase in spectral count. For example, the spectral counts of pregnancy zone protein are 2186 vs 208 in dPC(+) vs dPC(−). The detailed list of each identified peptides for the selected proteins and their corresponding hit values are reported in Table S-1. In addition, it is found that the number of miscleaved peptides were much smaller in the dPC(+) than in the dPC(−) approach (approximately a 2−3 fold reduction), which may lower the false positive rate in protein identification. A possible explanation is that the reduced proteome complexity resulting from the dPC separation step could allow an improved performance for the trypsin digestion step.

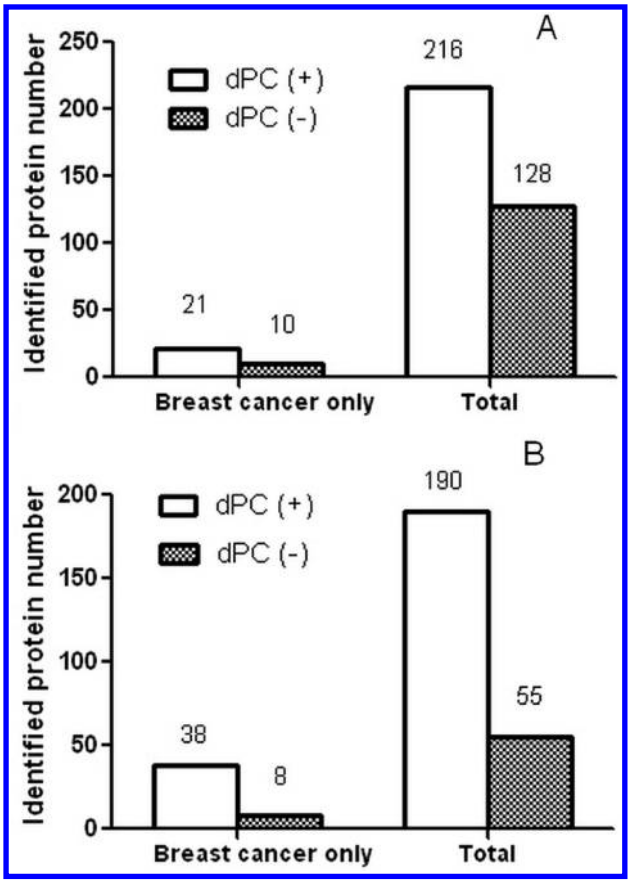


Figure 2. Number of proteins identified in the M-LAC (A) bound and (B) unbound fractions with and without IEF further separation. This analysis is based on proteins with more than one unique peptide.

Demonstration of Deeper Mining of the Proteome/Glycoproteome using Combined M-LAC and dPC Fractionations. The value of a combination of the first two steps, namely abundance protein depletion and M-LAC has been reported earlier.[10] The improved performance of the dPC(+) system

Table 2. Proteins Identified with 2-3 Unique Peptides (Spectral Count >1) through Depletion-(M-LAC)-dPC(+)-MS Approach

Protein description	Swiss-Prot acc. no.	SL[a]	spectral count[b] unique	spectral count[b] total	ΔPPM[c]	score[c]
Annexin A1	P04083	N, C, M	2	2	0.72	67
14-3-3 protein epsilon	P62258	C	2	2	1.18	61
Beta-actin-like protein 2	Q562R1	C	2	2	2.93	41
A disintegrin and metalloproteinase with thrombospondin motifs 13	Q76LX8	S	2	2	1.94	67
Creatine kinase M-type	P06732	C	2	2	0.70	79
Integrin beta-1	P05556	M	2	2	0.86	28
Tropomyosin beta chain	P07951	C	2	2	1.88	45
Caspase-14	P31944	C, N	2	2	3.35	34
Structural maintenance of chromosomes protein 3	Q9UQE7	N	2	3	2.43	30
4F2 cell-surface antigen heavy chain	P08195	M	2	5	3.31	42
SH3 domain-binding glutamic acid-rich-like protein 3	Q9H299	C, N	2	6	4.14	57
Filaggrin-2	Q5D862		2	7	1.46	49
Tenascin-X	P22105	S	5	9	1.47	47
Serum amyloid A-4 protein	P35542	S	2	11	2.35	48
Phosphoglycerate kinase 2	P07205	C	2	13	2.85	31
Selenoprotein P	P49908	S	2	22	2.35	61
ATP synthase subunit beta, mitochondrial	P06576	PM	2	19	2.37	59
Alpha-actinin-2	P35609	C	3	3	3.21	49
Alpha-mannosidase 2	Q16706	G, M	3	3	1.84	37
Gamma-glutamyl hydrolase	Q92820	S	3	4	3.07	42
Contactin-1	Q12860	M	3	4	2.29	48
Di-*N*-acetylchitobiase	Q01459	L	3	5	2.88	59
Neural cell adhesion molecule L1-like protein	O00533	M	3	7	2.61	34
Xaa-Pro dipeptidase	P12955		3	8	3.43	53
Basement membrane-specific heparan sulfate proteoglycan core protein	P98160	S, M	3	8	2.79	38
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	P33908	G, M	3	9	2.60	46
Putative HLA class I histocompatibility antigen, alpha chain H	P01893	M	3	14	1.92	39
Serum amyloid A protein	P02735	S	3	18	3.31	47

[a] SL: subcellular location as annotated in Swiss-Prot.[20] C = Cytoplasm; M = Membrane; G = Golgi; PM = Peripheral membrane; CS = cell surface; N = Nucleus; ER = Endoplasmic reticulum; S = Secreted; L = Lysosome. [b] The sum of spectral count from two independent mass spec observations. [c] Average ΔPPM and MASCOT Score values are shown for each protein.

which includes all three fractionation steps is shown in Figure 2. The total protein identification numbers are 216 vs 128 and 190 vs 55 as dPC(+) vs dPC(−) in M-LAC bound (Figure 2A) and unbound (Figure 2B) fractions, respectively. In addition, 11 and 30 additional proteins unique to the breast cancer proteome were identified with our dPC (+) approach in the bound and unbound fraction, respectively.

As is shown in Table 2 and Table S-2-1 (Supporting Information), a significant number of lower level proteins were identified in the serum samples collected from breast cancer and healthy individuals and these proteins were not detected in the absence of the dPC step. Lower level proteins that were identified with 2−3 unique peptides and spectral counts of greater than 2 are shown in Table 2 (also see Table S-2-1 for proteins identified with 4−10 unique peptides, as well as the identified peptides for each protein in detail, Supporting Information). Since we used stringent criteria ($p < 0.005$) to filter peptides after the database search (see the Experimental section), the identified peptides are with relatively high mass accuracy (ΔPPM is 2.44 on average). Hence, we have also reported on proteins with one peptide identification in Table S-2-2 (Supporting Information). In addition, the MS2 spectra of these peptides have been examined manually to ensure a confident identification. Further evidence for the deeper mining ability of the dPC(+) platform is demonstrated by the detection of the low abundance secreted proteins and potential tissue leakage fragments. According to protein annotations in Swiss-Prot,[20] the subcellular locations of proteins in Tables 2 and S-2-2 (Supporting Information) include secreted, membrane, cytoplasm, Golgi, peripheral membrane, cell surface, nucleus etc. The presence of proteins in serum samples which have a non secreted classification can be explained by mechanisms, such as alternatively spliced forms, the cleavage of plasma membrane associated proteins by proteases[21] and damage-associated molecular pattern (DAMP) in cell turn over.[22] Secreted proteins account for 34% and 15% as shown in Table S-2-1 (proteins with 2−10 unique peptides) and Table S-2-2 (proteins with 1 total peptide), respectively. In addition, some of the secreted proteins listed in Tables 2 and S-2-2 have been studied using ELISA by other laboratories and demonstrated to be of low abundance. For example, periostin, tenascin-X and vascular cell adhesion protein 1 have been reported to be 21.0 ± 7.3 ng/mL,[23] 476 ± 117 ng/mL[17] and 574.2 ± 42.7 ng/mL.[24] In the future study, the detection range could be further extended by depleting more high abundance

Table 3. Protein of Interest Identified through the Depletion-(M-LAC)-dPC(+)-MS Approach

Protein description	Swiss-Prot acc. no.	ΔM-LAC[a]	Δ*p*I[a]	Δabundance[a]	significance in cancer[b]
4F2 cell-surface antigen heavy chain	P08195		−		+
72 kDa type IV collagenase	P08253		−		+
Actin, cytoplasmic 1	P60709			+	+
Adiponectin	Q15848		−		++
Alpha-1-acid glycoprotein 2	P19652			+	++
Alpha-1-antichymotrypsin	P01011	−	−		++
Alpha-1-antitrypsin	P01009			+	++
Alpha-1B-glycoprotein[c,d]	P04217		−		++
Alpha-2-antiplasmin	P08697	−	−		++
Alpha-2-HS-glycoprotein	P02765		−		++
Angiotensinogen	P01019	−	−		++
Antithrombin-III	P01008		−		++
Apolipoprotein B-100	P04114	−			+
Biotinidase	P43251		−		++
C4b-binding protein alpha chain	P04003		−		+
Cadherin-5	P33151		−		++
Carbonic anhydrase 1	P00915		+		+
Cartilage oligomeric matrix protein (Thrombospondin-5, TSP5)[c,e]	P49747			+	++
CD5 antigen-like	O43866	−		+	+
Cholinesterase	P06276		−		++
Clusterin	P10909	−	−		++
Collagen alpha-1(I) chain	P02452			+	++
Complement C1q subcomponent subunit B	P02746			−	++
Complement C1s subcomponent	P09871	−			++
Complement C4-A	P0C0L4			+	++
Complement C4−B	P0C0L5	−			++
Desmoglein-2	Q14126		−		++
Desmoplakin	P15924			+	++
Di-*N*-acetylchitobiase	Q01459		−		++
Extracellular matrix protein 1	Q16610		−		++
Glutathione peroxidase 3	P22352		−	+	+
Haptoglobin	P00738		−		++
Hemopexin	P02790		−		++
Heparin cofactor 2	P05546		−		+
Hepatocyte growth factor-like protein	P26927			−	+
Histidine-rich glycoprotein	P04196	−	−		++
Hyaluronan-binding protein 2	Q14520		−		+
Immunoglobulin J chain	P01591			+	+
Insulin-like growth factor-binding protein complex acid labile subunit	P35858		−		+
Interalpha-trypsin inhibitor heavy chain H1	P19827	−	−		++
Interalpha-trypsin inhibitor heavy chain H2	P19823		−		+
Interalpha-trypsin inhibitor heavy chain H3	Q06033	−			++
Interalpha-trypsin inhibitor heavy chain H4	Q14624	−	−		++
Kallistatin	P29622		−		++
Kininogen-1	P01042	−	−		++
Lactotransferrin	P02788		−	+	++
Lumican	P51884		−		++
Mannan-binding lectin serine protease 1	P48740		−		+
Monocyte differentiation antigen CD14	P08571	−	−		++
Myosin-1	P12882			−	+
Myosin-2	Q9UGM7			−	+
Myosin-4	Q9UGM8			−	++
N-Acetylmuramoyl-L-alanine amidase	Q96PD5		−		+
Neural cell adhesion molecule L1-like protein	O00533		−		++

dx.doi.org/10.1021/ac2002802 |*Anal. Chem.* 2011, 83, 4845–4854

Table 3. Continued

Protein description	Swiss-Prot acc. no.	ΔM-LAC[a]	Δp*I*[a]	Δabundance[a]	significance in cancer[b]
Peroxiredoxin-2	P32119			+	++
Phosphatidylcholine-sterol acyltransferase	P04180		−		+
Phospholipid transfer protein	P55058		−		++
Pigment epithelium-derived factor	P36955		−		++
Plasma protease C1 inhibitor	P05155		−	−	++
Plasma serine protease inhibitor	P05154		−		++
Plasminogen	P00747	+			++
Plastin-2	P13796			+	+
Polymeric immunoglobulin receptor	P01833		−		++
Pregnancy zone protein	P20742	−			++
Properdin	P27918		−		++
Protein AMBP	P02760	−	−		+
Protein Z-dependent protease inhibitor	Q9UGM6		−		++
Prothrombin	P00734	−			++
Retinol-binding protein 4	P02753	−	−	+	
Scavenger receptor cysteine-rich type 1 protein M130	Q86VB7		−		++
Serotransferrin	P02787	+			++
Serum amyloid A protein	P02735		−		++
Serum amyloid P-component[c,f]	P02743		−	+	++
Sex hormone-binding globulin	P04278		−		++
Tenascin-X[c,g]	P22105			+	++
Thrombospondin-1 (TSP1)[c,e,h]	P07996	−	+		++
Thyroxine-binding globulin	P05543		−		+
Transthyretin	P02766			+	+
Vascular cell adhesion protein 1	P19320		−		++
Vitamin K-dependent protein C	P04070	−	−		
Vitamin K-dependent protein S	P07225	−	−		
Vitronectin	P04004		−		+
Von Willebrand factor	P04275		−		++
Zinc-alpha-2-glycoprotein	P25311		−		++

[a] AM-LAC: altered binding to M-LAC; Δp*I*: p*I* change in calculated pI vs measured pI or in dPC profile breast cancer vs control samples; ΔAbundance: abundance change. "±" refers to "increase/decrease" level. The criteria used to define an increase or decrease in level is described in the Supporting Information. [b] Biological significance in (breast) cancer is based on the listed references. "+/++" indicates that the protein has been reported to be associated with "cancer/breast cancer", respectively. References are in Supporting Information Table S-4. [c] Five proteins underlined will be discussed in detail in "Proteins of interest showing abundance change" in the "Results and discussion" section. [d] Was reported to exhibit a p*I* shift in breast cancer serum sample.[7] [e] TSP1 and TSP5 interact with cell surface proteins such as integrins and CD47 and glycosaminoglycans. They may play important roles in cancer progression by enhancing the adhesion and migration of tumor cells and stromal cells. [26] [f] SAP is not a significant acute-phase protein even though it has substantial homology with C-reactive protein (CRP).[27] [g] An extracellular matrix glycoprotein, which regulates cell adhesion, growth and cellular migration.[28] [h] TSP1 is an adhesive glycoprotein that has multiple biological functions, including platelet aggregation, fibrin clot formation, cell−cell/matrix contact, activation of extracellular proteases and inhibition of angiogenesis.[29]

protein, or by pooling less dPC gel plugs to yield more fractions for MS analysis.

Selection of Proteins of Potential Interest for Breast Cancer Detection. One of the challenges of a clinical proteomic study is to effectively select protein candidates for further study because downstream validation studies require expensive and time-consuming developments, such as ELISA and the multiple reaction monitoring (MRM) approach,[25] on a cohort of clinical samples which are usually difficult to obtain. In this study of the breast cancer serum proteome with the dPC(+) approach, a total of approximately 400 proteins were characterized and four criteria of interest were investigated for each identified protein, namely M-LAC binding, potential p*I* shift, abundance change and biological/clinical significance (detailed information is shown in Table S-3-1, S-3-2 and S-3-3, Supporting Information). Table 3 lists a subset of ∼80 proteins from the total of 400 with an observation in at least two of the four categories. Clearly, this number of candidates is still too large for a follow up study and thus, we performed a further selection based on the requirement for a significant and positive observation in at least two categories, which includes the data present in Table S-3-1 to S-3-3, and the protein's cancer/biological significance (also see Table 4). As example of this selection process, we report here on five proteins that meet these criteria: thrombospondin-1 and 5 (TSP1 and TSP5), alpha-1B-glycoprotein (A1BG), serum amyloid P-component (SAP) and tenascin-X (TN-X). The improved detection limit over comparable approaches and the revealed potential changes (abundance or/and glycan) demonstrate the advantages of our platform including extending the dynamic range of the proteomic measurement, as well as providing information on glycan pattern and p*I* profile.

Proteins of Interest Showing Abundance Change. We will now describe the salient features of these 5 proteins selected for

Table 4. Spectral Count of Five Protein Candidates and Peak Area of Selected Peptides in Control and Breast Cancer Sera

Protein Description	Swiss-Prot acc. no.	M-LAC fractions	Spectral count [a] Con [b]	Spectral count [a] BC [b]	IEF fraction # [c]	PA 1 [d] Con	PA 1 [d] BC	PA 2 [e] Con	PA 2 [e] BC
Abundance change									
Serum amyloid P-component (SAP)	P02743	UNB+BD [f]	3	68↑↑ [g]	3	ND [h]	5.50E+08↑↑	ND	4.53E+08
					4	ND	1.01E+09↑↑	3.63E+07	7.34E+08
Thrombospondin-5 (TSP5)	P49747	UNB+BD	ND	20↑↑	1	ND	1.06E+06↑↑	ND	1.72E+07
					2	ND	1.40E+08↑↑	ND	2.90E+07
pI shift									
Alpha-1B-glycoprotein (A1BG)	P04217	UNB+BD	319	398↑	2	7.57E+08	1.21E+09	2.35E+09	2.92E+09
					3	1.26E+09	1.16E+09	2.69E+09	1.60E+09
Different M-LAC binding									
Thrombospondin-1 (TSP1)	P07996	UNB	1	19↑↑	4	ND	5.61E+06↑↑	ND	4.46E+06
		BD	147	120↓	4	3.09E+07	2.51E+07↓	2.62E+07	2.00E+07
Different M-LAC binding, pI shift and abundance change									
Tenascin-X (TN-X) [i]	P22105	UNB	3	3	4	1.73E+06	1.35E+06↓		
		BD	3	6↑	1 and 2	1.90E+06	2.48E+06↑	(1.31 fold change)	

[a] The sum of spectral count from two independent mass spec observations. [b] Con: control; BC: breast cancer. [c] IEF fractions with the highest concentration of each protein among the 10 IEF fractions. For example, dPC fraction #4 and #3 are the most and the second most focused fractions for SAP, respectively. [d] Peak area of the selected peptide for each protein. Namely, peptide DNELLVYK, peptide NALWHTGDTESQVR, peptide HQFLLTGDTQGR, peptide GGVNDNFQGVLQNVR and peptide LNWEAPPGAFDSFLLR identified for SAP, TSP5, A1BG, TSP1 and TN-X, respectively. [e] Peak area of another selected peptide for each protein. Namely, peptide IVLGQEQDSYGGK, peptide QVCTDINECETGQHNCVPNSV-CINTR, peptide TPGAAANLELIFVGPQHAGNYR, peptide GTSQNDPNWVVR and peptide LGPISADSTTAPLEK identified for SAP, TSP5, A1BG, TSP1 and TN-X, respectively. [f] UNB: M-LAC unbound fraction; BD: M-LAC bound fraction. [g] ↑↑ and ↑/↓ refer to significant increased and slight increased/decreased level of the protein or peptide, respectively, when compare breast cancer to control case. [h] ND means Not Detected. [i] Only one peptide of TN-X from the IEF separation was available for good quality peak area comparison.

further study. Serum amyloid P-component (SAP) and thrombospondin-5 (TSP5 or cartilage oligomeric matrix protein) were identified with 6 and 7 unique peptides, respectively. As shown in Table 4, the total level of serum amyloid P-component (SAP) and thrombospondin-5 (TSP5) were increased in the breast cancer serum samples relative to controls based on the spectral count measurement (in fact, TSP5 was only detected in breast cancer serum samples). We then confirmed this observation with the corresponding peak area measurements where we selected two peptides for each protein based on the criteria of a consistent IEF profiles in all samples analyzed and good S/N ratios in the MS measurement (see Table 4 footnote c and d). Our results are also consistent with other cancer studies: the serum level of SAP has been reported to increase significantly in breast cancer and its concentration is correlated with disease stages;[30] elastosis with localized SAP has been observed at a higher level in breast carcinoma;[31] the thrombospondin family including TSP5 has also been reported to increase in malignant breast tissue.[32]

Proteins of Interest Showing Potential p*I* Shift. While a significant number of proteins listed in Table S-3-2 (73 proteins of interest showing p*I* shift, Supporting Information) showed a difference in measured vs theoretical p*I* of greater than 0.5 pH units, we have selected alpha-1B-glycoprotein (A1BG), a known glycoprotein, as a representative example. We observed a pI shift to a more acidic value for A1BG in the breast cancer samples (from fraction 3 to 2, see the peak area values with underlines in Table 4) but with only a slight increase in total abundance in the cancer samples. Differences in the glycan structure of A1BG has been reported to be associated with breast cancer,[33] pancreatic cancer[4] and colorectal cancer.[34]

Proteins of Interest Showing Different M-LAC Binding. With thrombospondin-1 (TSP1), we have an example where there is no concentration change (spectral count 139 vs 148, breast cancer vs control) but a difference in M-LAC binding where less TSP1 is bound to the M-LAC column in breast cancer (see Table 4, in addition, the full data set is given in Table S-3-1, Supporting Information). Such a result can be explained by a

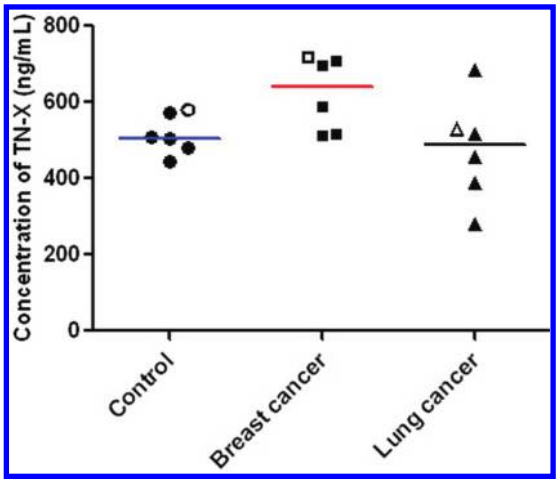


Figure 3. Levels of tenascin-X (TN-X) evaluated by ELISA. Sera from five control (●), five breast cancer (■) and five lung cancer (▲) individuals, as well as the pool of each set (○, □ and △) were studied.

change in glycan structure. This observation is consistent with the facts that TSP1 is known to be glycosylated and an abnormal glycan structure of TSP1 has been reported in congenital disorders of glycosylation (CDGs).[35] Moreover, the association of TSP1 with breast cancer has been shown in other publications.[36−39]

Proteins of Interest Showing Different M-LAC Binding, p*I* Shift and Abundance Change. Finally, we have selected an example, tenascin-X (TN-X), from these studies where there is evidence for changes in concentration and glycan structure, as well as relevant disease information. Increased levels of tenascin-X in tumor development[40−42] or, more specifically, breast cancer[43,44] has been reported. Although TN-X is a relatively medium level serum protein (476 ± 117 ng/mL),[17] we observed an increased serum level of tenascin-X (TN-X) in the breast cancer pooled sample from the spectral count measurement, as well as peak area measurements of the identified peptides (see Table 4). Furthermore, a potentially altered glycan structure of TN-X could be demonstrated both by different p*I* profiles and differential M-LAC binding. In the dPC analysis of the unbound and bound M-LAC fractions, TN-X was primarily focused in fraction 4 and (1 + 2), respectively. The level of spectral counts (3−6) was not sufficient for meaningful relative concentration measurement, but the peak area comparison did demonstrate a modestly increased level of TN-X in breast cancer serum sample (see Table 4).

Confirmation of Proteomic Analysis of Tenascin-X (TN-X). The next stage in a proteomic biomarker discovery, such as described here, is to develop a high throughput assay suitable for the clinical laboratory. At present, ELISA is the gold standard in terms of sensitivity, reproducibility and throughput.[45] The ELISA described here is used to confirm the presence of this protein in the serum samples and to compare the M-LAC, dPC and MS results with an ELISA format. Hence, we performed an ELISA for tenascin-X in small study using 5 controls and 5 breast cancer patients. In addition, sera from 5 lung cancer patients were included for comparative purposes. As shown in Figure 3, the mean level of tenascin-X is somewhat increased for the ELISA study in breast cancer compared to the control group with a fold change of 1.4, which is consistent with the 1.3 fold change (see Table 4) measured by our proteomic analysis. In the case of lung cancer, the levels of tenascin-X were not increased, which is consistent with the absence of literature reports. This small study indicates that this antibody can be used as an assay reagent for tenascin-X, as well as confirming the proteomic results of a somewhat increased level of this protein in breast cancer. As we discussed in Table 3, a protein disease marker may exhibit changes either in concentration, glycosylation, or both. Thus, the ELISA format which is typically based on a protein epitope may not be sufficient. These initial results indicate the need for a glycoproteomics measurement for a more discriminating assay, which is based on glycan as well as protein concentration changes. As previously published,[7] we will use the lectin array format developed by Dr. Haab to measure the glycan changes in tenascin-X in a larger sample set of patients with breast cancer.

■ CONCLUSIONS

In this study, we report on the development of a multidimensional fractionation platform including high abundance protein depletion, multi-lectin chromatography (M-LAC) separation, IEF (dPC) fractionation, and LC−MS/MS analysis for the investigation of the breast cancer serum (glyco)proteome. We have demonstrated that the dPC(+) platform significantly extends the dynamic range of the serum proteomic measurement. As a result, we have identified more low abundance proteins with higher sequence coverage relative to the approach without dPC fractionation. We have also given examples of another advantage of this platform, in that proteins with altered glycans can be observed to change their binding affinity toward the multi-lectin column (neutral and charged glycans), as well as showing different p*I* profiles in IEF (dPC) (due to the presence of different amounts of sialic acid). We plan to follow the characterization study reported here, with assays suitable for larger clinical studies, such as ELISA, MRM of the corresponding glycopeptide and an antibody-lectin sandwich microarray to confirm these glycoprotein biomarker candidates in a significant group of patients in a follow up study. The steps will be performed in a manner consistent with the goal of discovery of biomarkers for the improved early detection for breast cancer.

■ ASSOCIATED CONTENT

Supporting Information. Supplemental experimental information, figures, and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: m.hincapie@neu.edu. Fax: 617-373-2855.

Present Address

⊥Ventana Medical Systems, Inc. Oro Valley, Arizona, United States.

■ ACKNOWLEDGMENT

This study was supported by National Cancer Institute grant U01-CA128427 and Korean Research WCU grant R31-2008-000-10086-0. We are grateful to Dr. Frank Detmers and BAC. B.V. for providing the CaptureSelect depletion resins. We also thank Dr. Shiaw-Lin Wu and Dr. Tomas Rejtar for advice on this project and Majlinda Kullolli for technical support. This is Contribution Number 980 from the Barnett Institute.

REFERENCES

(1) *Cancer Facts and Figures*; American Cancer Society: Atlanta, 2010.

(2) Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J.-C. *Electrophoresis* **2000**, *21*, 1104–1115.

(3) Bosques, C. J.; Raguram, S.; Sasisekharan, R. *Nat. Biotechnol.* **2006**, *24*, 1100–1101.

(4) Zhao, J.; Qiu, W.; Simeone, D. M.; Lubman, D. M. *J. Proteome Res.* **2007**, *6*, 1126–1138.

(5) Zhao, J.; Patwa, T. H.; Qiu, W.; Shedden, K.; Hinderer, R.; Misek, D. E.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M. *J. Proteome Res.* **2007**, *6*, 1864–1874.

(6) Haab, B. B. *Mol. Cell. Proteomics* **2005**, *4*, 377–383.

(7) Zeng, Z.; Hincapie, M.; Haab, B. B.; Hanash, S.; Pitteri, S. J.; Kluck, S.; Hogan, J. M.; Kennedy, J.; Hancock, W. S. *J. Chromatogr., A* **2010**, *1217*, 3307–3315.

(8) Yang, Z.; Hancock, W. S. *J. Chromatogr., A* **2004**, *1053*, 79–88.

(9) Wang, Y.; Wu, S.-l.; Hancock, W. S. *Glycobiology* **2006**, *16*, 514–523.

(10) Kullolli, M.; Hancock, W. S.; Hincapie, M. *Anal. Chem.* **2009**, *82*, 115–120.

(11) Gianazza, E. *J. Chromatogr., A* **1995**, *705*, 67–87.

(12) Tang, H.-Y.; Ali-Khan, N.; Echan, Lynn A.; Levenkova, N.; Rux, John J.; Speicher, David W. *PROTEOMICS—Clin. Appl.* **2005**, *5*, 3329–3342.

(13) Wehr, T. *LC-GC North Am.* **2008**, *26*, 930–936.

(14) Zilberstein, G.; Korol, L.; Bukshpan, S.; Baskin, E. *PROTEOMICS—Clin. Appl.* **2004**, *4*, 2533–2540.

(15) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

(16) Egging, D. F.; Peeters, A. C. T. M.; Grebenchtchikov, N.; Geurts-Moespot, A.; Sweep, C. G. J.; Den Heijer, M.; Schalkwijk, J. *FEBS J.* **2007**, *274*, 1280–1289.

(17) Schalkwijk, J.; Zweers, M. C.; Steijlen, P. M.; Dean, W. B.; Taylor, G.; van Vlijmen, I. M.; van Haren, B.; Miller, W. L.; Bristow, J. *N. Engl. J. Med.* **2001**, *345*, 1167–1175.

(18) Dayarathna, M. K. D. R.; Hancock, W. S.; Hincapie, M. *J. Sep. Sci.* **2008**, *31*, 1156–1166.

(19) Plavina, T.; Wakshull, E.; Hancock, W. S.; Hincapie, M. *J. Proteome Res.* **2006**, *6*, 662–671.

(20) Expert Protein Analysis System Proteomics Server. http://www.expasy.org (accessed Nov 2010).

(21) Butler, G. S.; Dean, R. A.; Smith, D.; Overall, C. M. *Methods Mol. Biol.* **2009**, *528*, 159–176.

(22) Green, D. R.; Ferguson, T.; Zitvogel, L.; Kroemer, G. *Nat. Rev. Immunol.* **2009**, *9*, 353–363.

(23) Ben, Q. W.; Zhao, Z.; Ge, S. F.; Zhou, J.; Yuan, F.; Yuan, Y. Z. *Int. J. Oncol.* **2009**, *34*, 821–828.

(24) Ohga, E.; Nagase, T.; Tomita, T.; Teramoto, S.; Matsuse, T.; Katayama, H.; Ouchi, Y. *J. Appl. Physiol.* **1999**, *87*, 10–14.

(25) Whiteaker, R. J.; Zhao, L.; Zhang, Y. H.; Feng, L.-C.; Piening, D. B.; Anderson, L.; Paulovich, G. A. *Anal. Biochem.* **2007**, *362*, 11.

(26) Kazerounian, S.; Yee, K. O.; Lawler, J. *Cell. Mol. Life Sci.* **2008**, *65*, 700–712.

(27) Skinner, M.; Vaitukaitis, J.; Cohen, A. S.; Benson, M. D. *J. Lab. Clin. Med.* **1979**, *94*, 633–638.

(28) Ikuta, T.; Ariga, H.; Matsumoto, K.-i. *Genes Cells* **2000**, *5*, 913–927.

(29) Adams, J. C.; Tucker, R. P. *Dev. Dyn.* **2000**, *218*, 280–299.

(30) Levo, Y.; Wollner, S.; Treves, A. J. *Scand. J. Immunol.* **1986**, *24*, 147–151.

(31) Mera, S. L.; Davies, J. D. *J. Pathol.* **1987**, *151*, 103–110.

(32) Pratt, D. A.; Miller, W. R.; Dawes, J. *Eur. J. Cancer Clin. Oncol.* **1989**, *25*, 343–350.

(33) Jung, K.; Cho, W.; Regnier, F. E. *J. Proteome Res.* **2008**, *8*, 643–650.

(34) Qiu, Y.; Patwa, T. H.; Xu, L.; Shedden, K.; Misek, D. E.; Tuck, M.; Jin, G.; Ruffin, M. T.; Turgeon, D. K.; Synal, S.; Bresalier, R.; Marcon, N.; Brenner, D. E.; Lubman, D. M. *J. Proteome Res.* **2008**, *7*, 1693–1703.

(35) Heinonen, T. Y. K.; Maki, M. *Am. J. Med.* **2009**, *41*, 2–10.

(36) John, A. S.; Rothman, V. L.; Tuszynski, G. P. *J. Oncol.* 2010, *2010*, No. 645376.

(37) John, A. S.; Hu, X.; Rothman, V. L.; Tuszynski, G. P. *Exp. Mol. Pathol.* **2009**, *87*, 184–188.

(38) Hyder, S. M.; Liang, Y.; Wu, J. *Int. J. Cancer* **2009**, *125*, 1045–1053.

(39) Wang, T. N.; Qian, X.-h.; Granick, M. S.; Solomon, M. P.; Rothman, V. L.; Berger, D. H.; Tuszynski, G. P. *J. Surg. Res.* **1996**, *63*, 39–43.

(40) Hasegawa, K.; Yoshida, T.; Matsumoto, K.-i.; Katsuta, K.; Waga, S.; Sakakura, T. *Acta Neuropathol.* **1997**, *93*, 431–437.

(41) Goepel, C.; Buchmann, J.; Schultka, R.; Koelbl, H. *Gynecol. Oncol.* **2000**, *79*, 372–378.

(42) Verstraeten, A. A.; Mackie, E. J.; Hageman, P. C.; Hilgers, J.; Schol, D. J.; Jongh, G. J.; Sshalkwjjk, J. *Brit. J. Dermatol.* **1992**, *127*, 571–574.

(43) Shoji, T.; Kamiya, T.; Tsubura, A.; Hatano, T.; Sakakura, T.; Yamamoto, M.; Morii, S. *Virchow Arch.* **1992**, *421*, 53–56.

(44) Ishihara, A.; Yoshida, T.; Tamaki, H.; Sakakura, T. *Clin. Cancer Res.* **1995**, *1*, 1035–1041.

(45) Zangar, R. C.; Daly, D. S.; White, A. M. *Expert Rev. Proteomics* **2006**, *3*, 37–44.