Proteome Profiling of Breast Tumors by Gel Electrophoresis and Nanoscale Electrospray Ionization Mass Spectrometry

**Louise Alldridge,[†] Gergana Metodieva,[‡] Christina Greenwood,[†] Khalid Al-Janabi,[§]
Laura Thwaites,[§] Paul Sauven,[†,||] and Metodi Metodiev*[,‡]**

*The Helen Rollason Research Laboratory, Helen Rollason Heal Cancer Charity,
Chelmsford, Essex CM1 1LL, United Kingdom, Department of Biological Sciences, University of Essex, Wivenhoe
Park, Colchester, Essex CO4 3SQ, United Kingdom, Histopathology Department, Broomfield Hospital,
Broomfield, Chelmsford CM1 7ET, United Kingdom, and The Breast Unit, Chelmsford and Essex Hospital, New
Writtle Street, Chelmsford, Essex CM2 0QH, United Kingdom*

We have conducted proteome-wide analysis of fresh surgery specimens derived from breast cancer patients, using an approach that integrates size-based intact protein fractionation, nanoscale liquid separation of peptides, electrospray ion trap mass spectrometry, and bioinformatics. Through this approach, we have acquired a large amount of peptide fragmentation spectra from size-resolved fractions of the proteomes of several breast tumors, tissue peripheral to the tumor, and samples from patients undergoing noncancer surgery. Label-free quantitation was used to generate protein abundance maps for each proteome and perform comparative analyses. The mass spectrometry data revealed distinct qualitative and quantitative patterns distinguishing the tumors from healthy tissue as well as differences between metastatic and non-metastatic human breast cancers including many established and potential novel candidate protein biomarkers. Selected proteins were evaluated by Western blotting using tumors grouped according to histological grade, size, and receptor expression but differing in nodal status. Immunohistochemical analysis of a wide panel of breast tumors was conducted to assess expression in different types of breast cancers and the cellular distribution of the candidate proteins. These experiments provided further insights and an independent validation of the data obtained by mass spectrometry and revealed the potential of this approach for establishing multimodal markers for early metastasis, therapy outcomes, prognosis, and diagnosis in the future.

Keywords: Breast cancer • SDS electrophoresis • mass spectrometry • Nanoelectrospray • Biomarkers

Introduction

Breast cancer accounts for 15% of all cancers diagnosed and for 30% of all female cancers in the U.K. (http://info.cancerresearchuk.org/cancerstats/incidence/commoncancers/, July 2006). Each year more than 42 000 people in the U.K. are diagnosed with the disease (99% of whom are women), affecting one in nine women throughout the course of their lifetime (http://info.cancerresearchuk.org/cancerstats/incidence/females/?a=5441, July 2006) . In the EU, breast cancer is now the second most commonly diagnosed cancer after bowel cancer, and is the most common cause of female deaths from cancer resulting in around 91 000 deaths per year (http://info.cancerresearchuk.org/cancerstats/geographic/cancerineu/incidenceandmortality/?a=5441, July 2006). Furthermore, the incidence of breast cancer has continued to rise in economically developed countries such as the U.K. for many years, placing an ever increasing burden upon public health services. While research has succeeded in highlighting several risk factors associated with breast cancer, many of the underlying causes remain largely unknown. Better markers are needed for diagnosis, prognosis, and optimization of therapy. Research into biomarkers using cell lines, human serum, and animal models has so far resulted in limited success due to the specificities of tumor biology. Proteomics is one of the new technologies promising to deliver breakthroughs into cancer biomarkers.[20] It should be especially efficient when applied directly to fresh specimens derived from human tumors. Its ability to grasp the qualitative and quantitative complexity of the proteome should allow the development of cancer biomarkers of a novel type—ones that are based on the expression patterns of multitude of proteins as opposed to single protein abundance. So far proteomics has been mostly used for the profiling of plasma and serum. It is tempting to use these readily obtainable materials, but a decade of serum and plasma proteomic investigations has shown that their proteomes are too complex to define using approaches relying on total

* Corresponding author: Dr. Metodi V. Metodiev, Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, U.K. Phone: +44-(0)-1206-873154. Fax: +44-(0)-1206-872592. E-mail: mmetod@essex.ac.uk.
† The Helen Rollason Research Laboratory.
‡ University of Essex.
§ Broomfield Hospital.
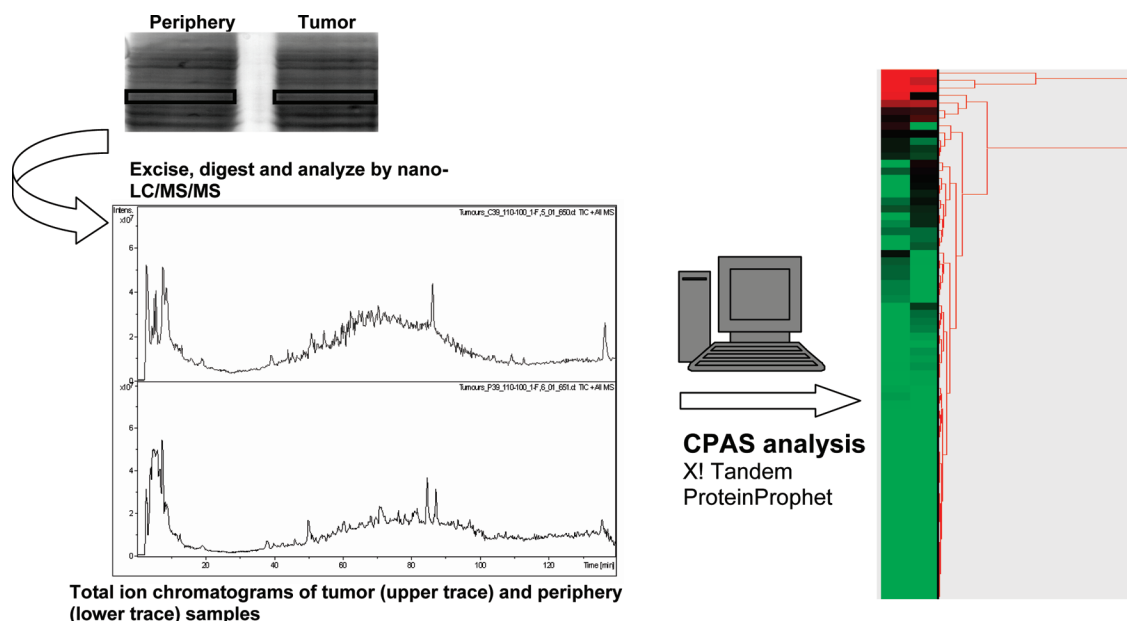|| Chelmsford and Essex Hospital.

Figure 1. Overview of the size-resolved differential profiling analysis of breast cancer specimens. The proteins of tumor and matching periphery tissue samples were separated by SDS-PAGE into 10 fractions and digested, and the resulting peptides were analyzed by nanoscale LC/MS/MS. Peptide fragmentation spectra were written into mzXML files and uploaded into the computational pipeline (CPAS) for analysis. Protein hit lists were filtered with ProteinProphet at 0.95 protein probability and exported to MS Excel files. The freeware program Hierarchical Clustering Explorer was used to produce the two-column heat map on the right showing protein abundances as estimated by spectral counting. The most abundant proteins are bright red, the least abundant, bright green.
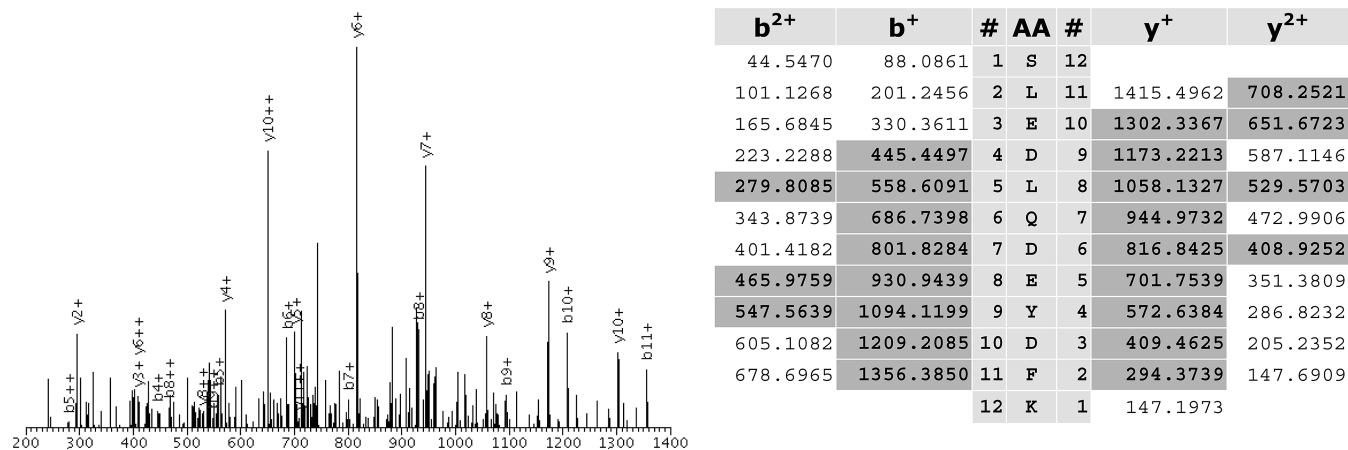

b^{2+}	b^+	#	AA	#	y^+	y^{2+}
44.5470	88.0861	1	S	12		
101.1268	201.2456	2	L	11	1415.4962	708.2521
165.6845	330.3611	3	E	10	1302.3367	651.6723
223.2288	445.4497	4	D	9	1173.2213	587.1146
279.8085	558.6091	5	L	8	1058.1327	529.5703
343.8739	686.7398	6	Q	7	944.9732	472.9906
401.4182	801.8284	7	D	6	816.8425	408.9252
465.9759	930.9439	8	E	5	701.7539	351.3809
547.5639	1094.1199	9	Y	4	572.6384	286.8232
605.1082	1209.2085	10	D	3	409.4625	205.2352
678.6965	1356.3850	11	F	2	294.3739	147.6909
		12	K	1	147.1973	

Figure 2. MS/MS spectrum of the tryptic peptide **SLEDLQDEYDFK** from STAT1 detected in the 100–110 kDa fraction of the proteome of one of the breast cancer tumor samples. Fragment assignments are shown in the table on the right-hand side by dark gray shading. The spectrum was acquired in a 150 min nano-LC/MS/MS experiment as described under Experimental Procedures. Fragment assignment, protein identification, and all statistical evaluations were performed automatically by the CPAS suite using X! Tandem and the PeptideProphet tool from Trans-proteomic Pipeline. B and Y fragment ion series are indicated above the corresponding ion peak in the spectrum.

protein identification by mass spectrometry. It is likely that the most useful molecules in unprocessed plasma and serum are present at concentrations below the detection limit of today's mass spectrometers. Fractionation and affinity enrichment can help, but they lead to an exponential increase of the number of samples to be analyzed, rendering the project unpractical. One alternative strategy is to look at the tumors themselves and identify the proteome patterns that distinguish them from healthy tissue or pathological phenotypes. If prospective markers overexpressed in tumor tissue are established, these can be assayed in serum and plasma using much more sensitive methods based on signal amplification such as Western blotting, ELISA, and antibody microarrays. The aim of this study was to explore the strategy of comparing the protein profiles of freshly collected breast cancer tumors and healthy tissue by combining established analytical methods such as 1D SDS-PAGE and low-flow electrospray ionization mass spectrometry supported by a suit of recently developed bioinformatics tools. This allows unattended analysis and integration of large-scale peptide fragmentation data sets. Validation experiments employing Western blotting and immunohistochemistry analyses of selected proteins showed that this approach has the potential to provide a powerful discovery platform in breast cancer biomarker and drug development research.

Table 1. Proteins Found To Be Strongly Overexpressed in Breast Tumors Compared to Corresponding Peripheral Tissue Samples ($n = 2$) and Healthy Mammary Gland Tissue Samples ($n = 2$)[a]

UniProt accession	protein name	protein function	slice (kDa)
P35579	Myosin-9	Motility	>150
Q5HY53	Filamin A	Actin binding, filament branching	>150
Q7Z351	Hypothetical protein	Antigen presentation	>150
Q9Y940	Talin-1	Attachment of the Actin cytoskeleton to the plasma membrane	>150
P46940	IQGAP1	Binds to activated CDC42 GAP but does not stimulate its GTPase activity	>150
O14746	Telomerase reverse transcriptase	Telomerase component	120–150
P01024	Complement C3	Activation of the complement system, inflamation	120–150
P18206	Vinculin	Cell adhesion, attachment of the Actin cytoskeleton to the plasma membrane	120–150
Q9Y4L1	150 kDa oxygen-regulated protein	Suppressor of hypoxia induced apoptosis, protein folding	120–150
Q13813	Spectrin alpha	Calcium-dependent movement of the cytoskeleton	120–150
Q8NF91	Nesprin	Nuclear envelop protein, Actin binding	120–150
Q9BSJ8	Fam62a	Membrane bound, contains C2 domains	120–150
P12814	Alpha actinin	Actin binding and bunling of F-Actin	100–110
P55072	TER ATPase	Transitional ER ATPase	100–110
P42224	STAT1	Signal transduction and activation of transcription	100–110
P20929	Nebulin	Binds to F-Actin and titin	100–110
P11021	GRP 78	Protein folding and complex assembly	70–100
P13667	Protein disulfide-isomerase A4	-S-S- rearrangement	70–100
P26038	Moesin	Connection of the cytoskeleton to the plasma membrane	70–100
P04220	Ig mu heavy chain	Antigen binding	70–100
P06396	Gelsolin	Actin deploymerization	70–100
P02787	Transferrin	Iron binding and transport	60–70
P13796	L-Plastin	Actin binding	60–70, 50–60
P19971	Thymidine phosphorylase	Catalyzes reversible phosphorolysis of thymidine. Promotes angiogenesis	50–60
P25705	ATP synthase alpha	ATP synthesis	50–60
P06576	ATP synthase beta	ATP synthesis	50–60
P30101	Protein disulfide isomerase A3	-S-S- rearrangement	50–60
Q15084	Protein disulfide isomerase A6	-S-S- rearrangement	50–60
P41219	Peripherin	Intermediate filaments	50–60
P00738	Haptoglobin precursor	Binds free hemoglobin	50–60
P78418	PEG3 protein	Zing finger protein, parentally imprinted	50–60
P01891	MHC class I antigen A-68	Antigen presentation	50–60
P10316	MHC class I antigen A-69	Antigen presentation	50–60
multiple	Immunoglobulin proteins	Antigen binding	50–60
Q562X4	Actin like protein	Protein binding	40–50
P13645	Keratin 10	Intermediate filaments	40–50
Q6NSB4	HP protein	Binds free hemoglobin	30–50
Q59HH1	Zinc finger protein 278	DNA binding	20–30
O00299	NCC27	Chloride ion channel	20–30
P67936	Tropomyosin 4	Actin binding	20–30
P52907	CapZ apha 1	Actin capping	20–30
P05388	L10E	Acidic ribosomal protein	20–30
P25786	Proteasome subunit alpha type 1	Protein turnover	20–30
Q7Z4W7	Rab like protein	GTP binding, protein transport, signal transduction	20–30
Q6JZS5	Nanog protein	Transcription factor, regulation of gene expression	20–30
P08572	Collagen alpha- 2(IV) chain	Component of the extracellular matrix	20–30
P55042	RAD protein	GTP binding, signal transduction	20–30
Q9BWA9	HNRPA2B1 protein	RNA binding	20–30
Q69YU5	Hypothetical protein DKFZp547P055	Unknown	10–20
Q06830	Peroxiredoxin-1	Redox regulation of the cell	10–20
P30041	Peroxiredoxin-6	Redox regulation of the cell	10–20
P62258	14–3–3 protein epsilon	Signal transduction	10–20
P02766	Transthyretin	Thyroid hormone binding	10–20
Q502W4	IGKC protein	Antigen presentation	10–20

[a] Two ER-/PgR-, node positive tumors were compared with their clinically normal periphery and with 2 normal breast tissue samples. Only proteins detected with more than 2 high-scoring peptides in both tumor samples but not detected in the peripheral or normal samples are included in the data set.

Experimental Procedures

Reagents. All chemicals, unless indicated otherwise in the text, were from Sigma-Aldrich. The highest available grades were used throughout the study. The primary antibodies were purchased from Abcam and Santa Cruz Biotechnology.

Fresh Tissue Collection. All tissue was collected under LREC and NHS Trust approval. All patients gave their informed consent for the use of their tissue, collected as part of routine treatment, in biochemical research. Surgical tissue was immediately placed on ice in the operating theater to ensure that the status of the tissue is as close as possible to its physiological

and biochemical status within the human body. The tissue was assessed and cut by the pathologist on ice before being divided up with a proportion of the tissue snap-frozen in liquid nitrogen and the remainder fixed in formalin with the shortest possible delay. Any tissue that was X-rayed or kept off ice was not used in this study. Breast tissue peripheral to the cancer (clinically normal) and normal breast tissue from patients undergoing noncancer breast surgery were collected and processed as described above.

Tumor Selection. For mass spectrometry analysis, tumors were compared directly with corresponding clinically normal

Table 2. Candidate Membrane Protein Markers for Early Metastatic Events[a]

UniProt accession	protein name	protein function	spectral count
Q6PYX1	Hepatitis B virus receptor binding protein	Receptor activity	68
P13796	L-Plastin	Motility	16
Q6N096	Unknown protein	Protein binding	15
P08133	Annexin 6	Ca binding	14
Q13965	CD36 antigen	Receptor	12
Q99536	VAT1	Membrane protein, quinone oxidoreductase	11
P12111	Collagen alpha-3(VI)	Cell binding	11
P26038	Moesin	Connects cytoskeleton to membrane	10
Q16853	Membrane copper amine oxidase	Cell adhesion	9
P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit	NADH oxidation	8
Q01518	CAP1, Adebylyl cyclase-associated protein 1	Filament dynamics	7
P07339	Cathepsin D	Protease, involved in breast cancer pathogenesis	6
P07384	Calpain 1	Protease, signaling	6
Q16891	Mitofilin	Cell proliferation inducing gene, mitochondrial inner membrane protein	6
Q562R1	Actin-like protein	Cytoskelleton	5
P62873	G beta protein	Signal transduction	5
P00505	Transaminase A	Fatty acid metabolism	5
Q7L4N5	Transmembrane protein 43	Unknown, multipass membrane protein	5
P49748	VLCAD	Fatty acid oxidation	5
P04217	Alpha 1B glycoprotein	Unknown, secreted protein	5
Q5TBT2	Fistronin	Actin binding	5
Q53HS5	Integrin beta 2	Cell adhesion	5
Q96QE2	Proton myo-inositol cotransporter	Multipass membrane protein, transporter	4
O75083	AIP1	Cytoskeleton organization	4
Q9Y490	Talin1	Cytoskeleton to plasma membrane contacts	4
O00299	CLC1	Chloride channel	4
Q16647	Prostacyclin synthase	Prostaglandin metabolism	3
Q2YD73	Coronin 1A	Cytoskeleton dynamics, phagocytosis	3
Q5VZ18	SHE	SH2 domain adapter protein	3

[a] Proteins identified with more than 2 peptides in the membrane fraction of pooled node negative cancer samples ($n = 3$) but not in the node-positive sample pool. The node-positive and node-negative tumors were matched for all other phenotypic parameters.

peripheral tissue, or were matched for size, type, histological grade, and receptor status (estrogen (ER), progesterone (PgR), and ErbB 2 (HER2) receptors) differing only in their nodal status. One group of tumors had nodes positive for cancer cells and the other was negative. This was to allow comparison of tumors that have displayed the ability to move from the primary site with those that have not to identify proteins that are related to the early stages of metastasis.

For initial validation of proteins selected from MS/MS data, a wide range of tumor samples were selected to assess consistency of expression in all types of breast cancers. For further analysis of proteins showing interesting trends, further tumor samples phenotypically matched but differing in nodal metastasis were chosen.

All tumor samples were identified so that the pathological findings could also be matched with prospective and retrospective clinical data from patient records and subsequent research findings.

Protein Sample Preparation. Proteins were extracted by mechanically homogenizing the samples in sample buffer containing 2% SDS, 50 mM Tris-HCl, pH 6.8, and protease and phosphatase inhibitors. Tissue homogenates were centrifuged at 16 000g for 15 min, and the clear supernatant was aliquoted and stored at -80 °C until needed for analysis. Prior to electrophoresis, the proteins were chloroform/methanol-precipitated and resuspended in the same sample buffer.

Membrane Protein Enrichment. Membrane proteins were enriched from tumor tissue using the Mem-PER eukaryotic membrane protein extraction kit from Pierce Biotechnology.

Gel Electrophoresis Prior to Mass Spectrometry. SDS-PAGE according to Laemmli[12] was performed with reduced samples using precast gradient minigels (Invitrogen) and standard protocols for separation and staining. Gels were stained with ethanol-based Coomassie Blue G-250 stain.

Protein Digestion and Preparation of Samples for Mass Spectrometry. Gel slices were subjected to in-gel digestion as previously described by Metodiev et al.[18] with some modifications to ensure better recovery of the peptides. Gel slices were destained by several changes of solution containing 100 mM NH_4HCO_3 and 50% acetonitrile and dehydrated in 100% acetonitrile before drying in a vacuum concentrator. Gel slices were reduced for 1 h in 12 mM DTT at 37 °C and subsequently alkylated with 55 mM iodacetamide in the dark. Gel slices were dehydrated again, dried, and rehydrated in trypsin solution at enzyme concentration of 20 μg/mL. Promega trypsin of MS grade quality was used throughout the experiments. After 18 h of incubation at 37 °C, the peptides were extracted and the samples evaporated to dryness. Evaporated peptide samples were stored at -80 °C until needed for analysis.

Nanoscale LC/MS/MS Analysis. The analysis of protein digests was performed using Ultimate II nano-LC (Dionex) coupled to Esquire HCT ion trap mass spectrometer (Bruker Daltonics) via online nano electrospray ion source. Sample introduction was via the micro autosampler Famos (Dionex). The peptides were separated by a biphasic 150 min binary gradient as follows: 10 min at 2% solvent B, followed by 90 min linear gradient to 30% B, followed by 40 min linear gradient to 50% B, followed by a wash for 5 min in 90 B. The flow was

Table 3. Candidate Membrane Protein Markers for Early Metastasis[a]

UniProt accession	protein name	protein function	spectral count
Q12907	VIP36	Secretion	8
Q08431	Lactadherin	Specific ligand for the alpha-v/beta-3 and alpha-v/beta-5 receptors	8
P08729	Cytokeratin 7	Cytoskeletal protein	8
Q9HAV0	GNB4	Signal transduction	6
Q6N090	Putative protein	Unknown	6
P20774 T	Mimecan	Induces bone formation in conjunction with TGF-beta-1 or TGF-beta-2	
:O60283	KIAA0534 protein	Sugar binding, development	5
P29043	Serpin-H1	Collagen binding protein, chaperon	5
Q07960	Rho GAP 1	Signal transduction	5
Q00325	Phosphate carrier protein	Ion transport	4
Q96AG4	Leucine-rich repeat-containing protein 59	Leucine rich membrane protein	4
Q01955	Collagen alpha-3(IV) chain	Structural protein	4
P59091	Unknown protein	Unknown	3
P43307	TRAP-alpha	Regulates Ca fluxes	3
P63244	Guanine nucleotide-binding protein subunit beta-2-like 1	Signal transduction	3
Q659E8	Putative protein	Oxidoreductase activity	3
Q8WZ27	Putative protein	Unknown	3
Q9NQC3	Nogo protein	Neurite outgrowth inhibitor	3
Q9HCK4	Roundabout homologue 2	Receptor for SLIT2	3
Q4VX75	Ezrin	Probably involved in connections of major cytoskeletal structures to the plasma membrane.	3
Q8TF66	Leucine-rich repeat-containing protein 15	Membrane protein	3
Q96SD1	Cytokeratin 6A	Cytoskeletal protein	3

[a] Proteins exclusively identified with more than 2 peptides in the membrane fraction of pooled node-positive cancer samples ($n = 3$) but not in the node-negative sample pool. The node-positive and node-negative tumors were matched for all other phenotypic parameters.

kept at 250–300 nL/min. The column used was Pepmap 75, a 75 μm i.d., 150 mm long reverse-phase column packed with 3 μm C18 silica particles (Dionex). The mass spectrometer was controlled by the manufacturer control software and operated in a data-dependent acquisition mode with the following settings: the threshold for detection was 25 000; the number of precursors was set to 3; the number of scans to be averaged was set to 3; the intelligent ion control (ICC) was set to 200 000. A mixed scan mode was used to select only multiply charged ions for MS/MS analysis. Under this mode, the instrument performs a survey scan at 8100 unit/s, which allows determination of charge states up to 3. Next, the 3 most intense multiply charged precursors are isolated and fragmented, and the fragments' m/z is determined using an ultra fast scan at 26 000 units/s. A dynamic exclusion was used to exclude analyzed precursors for 1 min after their second isolation by the instrument.

Bioinformatics and Data Analysis. MS/MS data were analyzed by an integrated analytical pipeline consisting of 2 computer workstations and a dedicated server using CPAS (Computational Proteomics Analysis System) and X! Tandem as search engine.[16] The raw MS data files were converted to mzXML files[21] using the compassXport program from Bruker Daltonics. The mzXML files were loaded into an automatic data analysis pipeline running under the experimental CPAS environment.[3,22] The CPAS suite version 1.4 was installed on a dual Xeon 64 bit processor server operating under the 64 bit SUSE 9.2 version of Linux and configured with Postgres 8.0.6 as a database and Jakarta Tomcat as a servlet container and web server. Postrgres and Tomcat were built from source prior to installation. The Transproteomic Pipeline (TPP) from the Seattle Institute of Systems Biology was integrated into the CPAS environment to provide statistical analysis of the protein identification data via its Peptide Prophet and Protein Prophet

tools.[11] Uniprot and the International Protein Index (IPI) databases were used to perform protein identification.

Gene Ontology Term Enrichment Analysis. Protein identified to be significantly overexpressed in the breast tumor tissue were analyzed using the GOLEM program[24] to identify biological process and molecular function categories enriched in the data set. The program uses the following formula for hypergeometric distribution to calculate the p-values for each category:

$$p = \sum_{j=k}^{n} \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}$$

where N is the total number of genes in the Human annotation file, M is the number of annotated genes in a given GO term, and n is the number of genes in the input data set. The p-value gives the probability that k or more genes from the data set will be assigned to a given GO term. Bonferoni correction for multiple hypotheses[2] was used as implemented in the GOLEM algorithm.

Quantitative Analysis. Relative abundance of the identified proteins was determined using the spectral counting method as described previously.[15] Comparison of protein profiles was achieved automatically by the "Compare Proteins" function of the MS2 module of CPAS. The results from the comparisons were exported to MS Excel files. The freeware program HCE (Hierarchical Clustering Explorer)[25] was used to integrate and visualize the protein abundance maps based on the quantitation by spectral counting.

Western Blotting Analysis. Breast cancer, tissue proximal to the tumor, and normal breast tissue were cut into small pieces while the tissue remained on dry ice and then homogenized in ice-cold lysis buffer containing 50 Mm NaCl, 1 mM EDTA, 10 Mm Tris-HCL, pH7.4, 1% Igepal, 1 mM EGTA, 1 mM
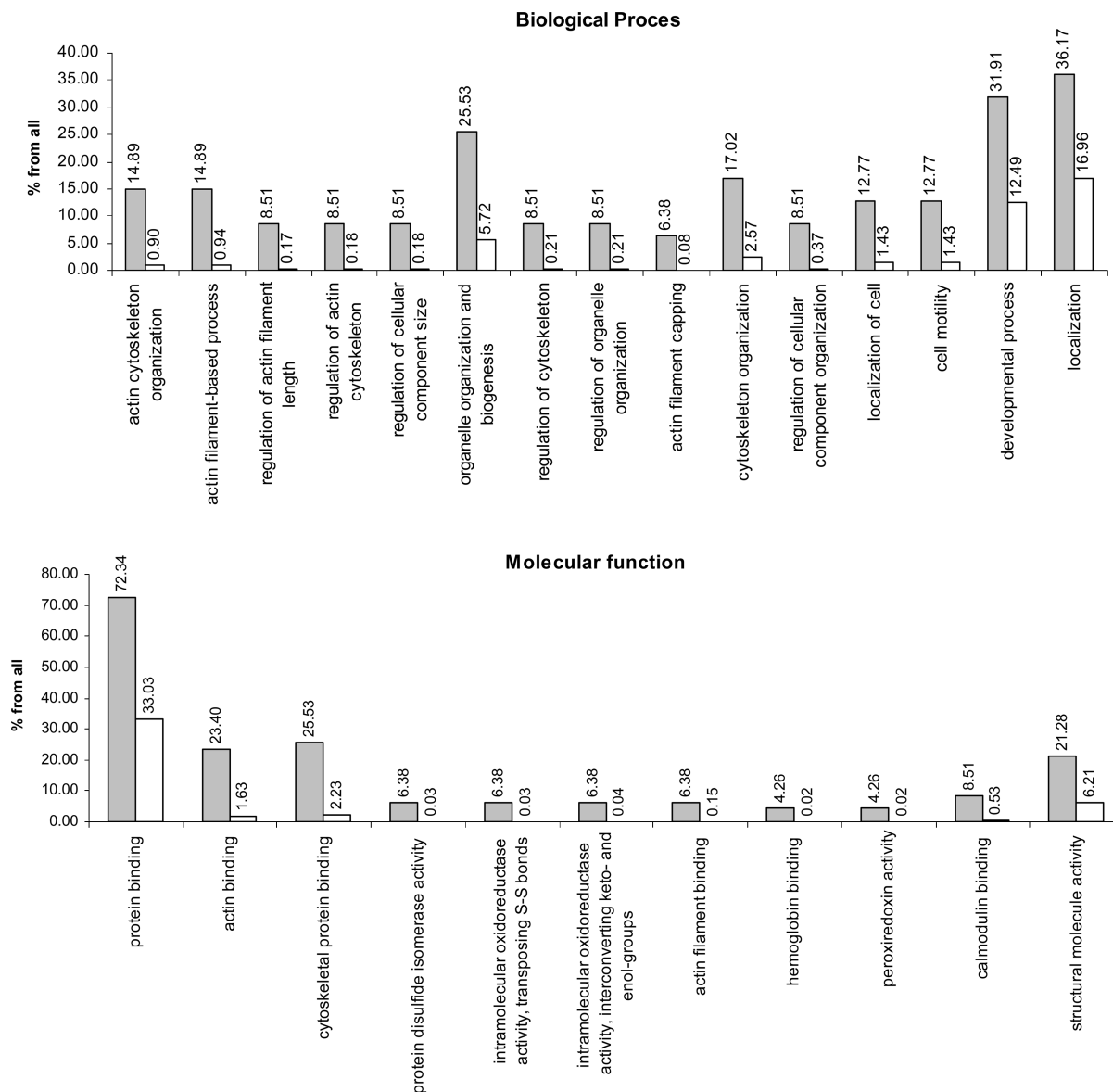
Biological Proces

(Bar chart, "% from all" axis. Values for gray/open bar pairs:)
- actin cytoskeleton organization: 14.89 / 0.90
- actin filament-based process: 14.89 / 0.94
- regulation of actin filament length: 8.51 / 0.17
- regulation of actin cytoskeleton: 8.51 / 0.18
- regulation of cellular component size: 8.51 / 0.18
- organelle organization and biogenesis: 25.53 / 5.72
- regulation of cytoskeleton: 8.51 / 0.21
- regulation of organelle organization: 8.51 / 0.21
- actin filament capping: 6.38 / 0.08
- cytoskeleton organization: 17.02 / 2.57
- regulation of cellular component organization: 8.51 / 0.37
- localization of cell: 12.77 / 1.43
- cell motility: 12.77 / 1.43
- developmental process: 31.91 / 12.49
- localization: 36.17 / 16.96

Molecular function

(Bar chart, "% from all" axis. Values for gray/open bar pairs:)
- protein binding: 72.34 / 33.03
- actin binding: 23.40 / 1.63
- cytoskeletal protein binding: 25.53 / 2.23
- protein disulfide isomerase activity: 6.38 / 0.03
- intramolecular oxidoreductase activity, transposing S-S bonds: 6.38 / 0.03
- intramolecular oxidoreductase activity, interconverting keto- and enol-groups: 6.38 / 0.04
- actin filament binding: 6.38 / 0.15
- hemoglobin binding: 4.26 / 0.02
- peroxiredoxin activity: 4.26 / 0.02
- calmodulin binding: 8.51 / 0.53
- structural molecule activity: 21.28 / 6.21

Figure 3. Identification of enriched Gene Ontology terms in the breast cancer data set from Table 1. Gene Ontology (GO) assignments for the identified proteins were from www.geneontology.org. The distribution of proteins across different molecular function and biological process types is represented by pairs of percentage bars. Open bars denote general Human proteome data, and gray bars are for proteins identified as significantly overrexpressed in breast carcinomas as compared to the matched peripheral tissue. GO categories enriched in the breast cancer data set were identified by calculating *p*-values using the hypergeometric distribution and the GOLEM program as described in Experimental Procedures. A *p*-value cutoff of 0.05 was used to generate the graphs; only GO terms with *p*-values lower than the cutoff are shown. The complete output of the GOLEM analysis is available in the Supporting Information.

PMSF, 1 μg/mL aprotinin, 1μg/mL leupeptin, 1 mM $Na_4P_2O_7$, and 25 mM NaF. Lysates were centrifuged to remove any fatty deposit, solid, and DNA/RNA contaminants. The protein concentration was determined by the Bio-Rad Detergent-Compatible protein assay.

Protein Separation by SDS Gel Electrophoresis and Western Blot Analysis. Protein samples were boiled in reducing sample buffer for 5 min. Proteins were resolved on acrylamide SDS gels and transferred to polyvinylidine difluoride membranes (Millipore).

Membranes were blocked with 3% (w/v) nonfat dried milk in PBS and incubated with the appropriate primary antibody at room temperature for 1 h. Following four washes in PBS-Tween 20, membranes were probed with the appropriate IR-conjugated secondary antibody (LI-COR Biosciences) at 1/2000 for a further hour at room temperature. Proteins were visualized using the ODYSSEY Infrared Imaging System (LI-COR Biosciences).

Proteins of interest were normalized against acitn expression and densitometric analysis was conducted using ODYSSEY software.

Immunohistochemistry. Immunohistochemistry was performed on a panel of paraffin-embedded sections of a range of breast tumor phenotypes, from tissue collected on ice as above. For antigen detection, deparaffinized sections were treated in a microwave for 20 min at 900 W in 10 Mm citrate buffer (pH 6.10). Endogenous peroxidase activity was inhibited with a mixture of H_2O_2/methanol (12 mL of H_2O_2 in 400 mL
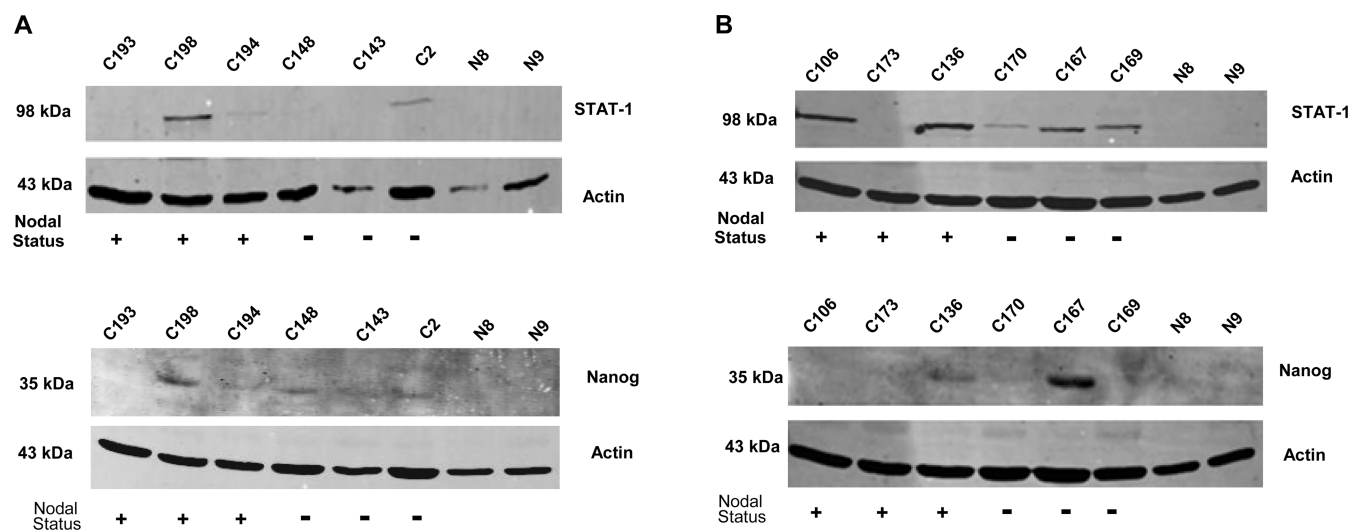
Figure 4. Stat1 and Nanog expression in carcinomas and normal breast tissue. Whole cell lysates from three individual cases of (A) ER+/PgR+ node positive (C193, C198, C194) and node negative (C148, C143, C2) Invasive Ductal Carcinomas (IDC [+/− DCIS]) and (B) ER-/PgR- node positive (C106, C173, C136) and node negative (C170, C167, C169) Invasive Ductal Carcinomas (IDC [+/− DCIS]) and tissue from two individual breast reduction mammoplasties (N8, N9) were subjected to SDS-PAGE followed by Western blot analysis. Total Stat1 and Nanog expression was detected using anti-Stat1 and anti-Nanog antibodies, respectively. Equal loading of lanes was maintained by performing a total protein assay and is confirmed by Western blot analysis using an anti-Actin antibody.

of methanol) for 12 min. Primary antibodies incubation was carried out for an hour at room temperature. Secondary biotinylated conjugated antibodies and tertiary ABC complexes (VectorLab) incubations were carried out for 30 min at room temperature. Detection of antibodies was performed using diamino-benzidine (DAB) as a chromogen (ImmPACT, Vector laboratories). Cell nuclei were counterstained with heamatoxylin. Slides were analyzed on an Olympus BX41 and images captured using image analysis software provided by Olympus.

Results and Discussion

Mass Spectrometry. In this study, we have analyzed fresh tumor, paired peripheral tissue, and normal tissue from the human mammary gland. We have also analyzed the membrane proteomes of pooled samples of 3 metastatic and 3 nonmetastatic breast cancers that were matched for all other phenotypic parameters. Each protein sample to be analyzed was separated into 10 size-resolved protein fractions by slicing the SDS-PAGE lanes into segments of 10–20 kDa. The slices were subjected to in-gel digestion and subsequent analysis by nano-LC/MS/MS. On average, 4–5000 searchable MS/MS spectra were recorded during the course of each individual LC/MS/MS run. The final raw data set consisted of more than 500 000 peptide fragmentation spectra. The raw data files were converted to mzXML and uploaded into CPAS for analysis. Figure 1 illustrates the experiment and the nature of the obtained results. For each analyzed sample, a heat map of protein abundances is generated. Clustering algorithms are then used to identify the proteins that are significantly overexpressed in tumor tissues compared to the periphery and the normal tissue. Similar heat maps can be used to compare different tumors or to compare tumors with cells grown in culture.

1. Size-Resolved Differential Profiling of Tumor versus Clinically Normal Peripheral Tissue Proteomes: Methodological Potential for the Detection of Biomarkers of Breast Cancer. More than 2000 proteins were positively identified based on ProteinProphet probability of 0.95 and above. About half of these proteins were identified with 2 or more peptides.

To ensure that candidate biomarker status is not assigned to occasional false-positive hits, we used an additional stringent filter: all proteins discussed in this study were identified by at least 2 peptides with PeptideProphet scores higher than 0.95 and CPAS Expect values lower than 0.01. A representative MS/MS spectrum demonstrating the quality of the acquired data and fragment assignments is shown in Figure 2. The spectrum is matched to a peptide derived from the transcription factor STAT1. STAT1 was identified with multiple high-score peptides in the 70–100 kDa fraction of tumor sample but was not detected in the periphery and normal samples. STAT1 is found in various tumors and human cancer cell lines; however, there is limited information on its role in breast cancer.[26] Lists of identified proteins with corresponding high-score peptides from 1 cancer and matched periphery are available in the Supporting Information.

The most abundant proteins in some of the fractions gave more than 200 spectral counts. Cytokeratin, albumin, and transferrin are three examples of very abundant proteins detected in the 60–70 kDa fractions of all analyzed samples from breast tissue.

Table 1 lists the proteins that were found to be significantly overexpressed in tumor tissue based on the analysis of 2 tumors and the corresponding peripheral tissue samples. Among these are several proteins potentially implicated in breast cancer such as L-plastin, STAT1, haptoglobin, and thymidine phosphorylase.[5,7,14,29] Proteins that are known to be involved in the regulation of the cytoskeleton are also prominent in this list. These include L-plastin, moesin, gelsolin, and talin. Chaperones such as GRP 78 and several protein disulfide isomerases were also found to be much more abundant in tumors than in the peripheral tissue and the healthy tissue.

Several proteins involved in signal transduction were detected in the tumor sample but not in the peripheral or healthy tissue. Such are IQGAP1, a 14–3–3 protein, and a Rab-like GTPase (Table 1). A 150 kDa oxygen regulated protein, a known cancer marker, was also detected only in the tumor tissue but not in the periphery or normal mammary gland tissue.
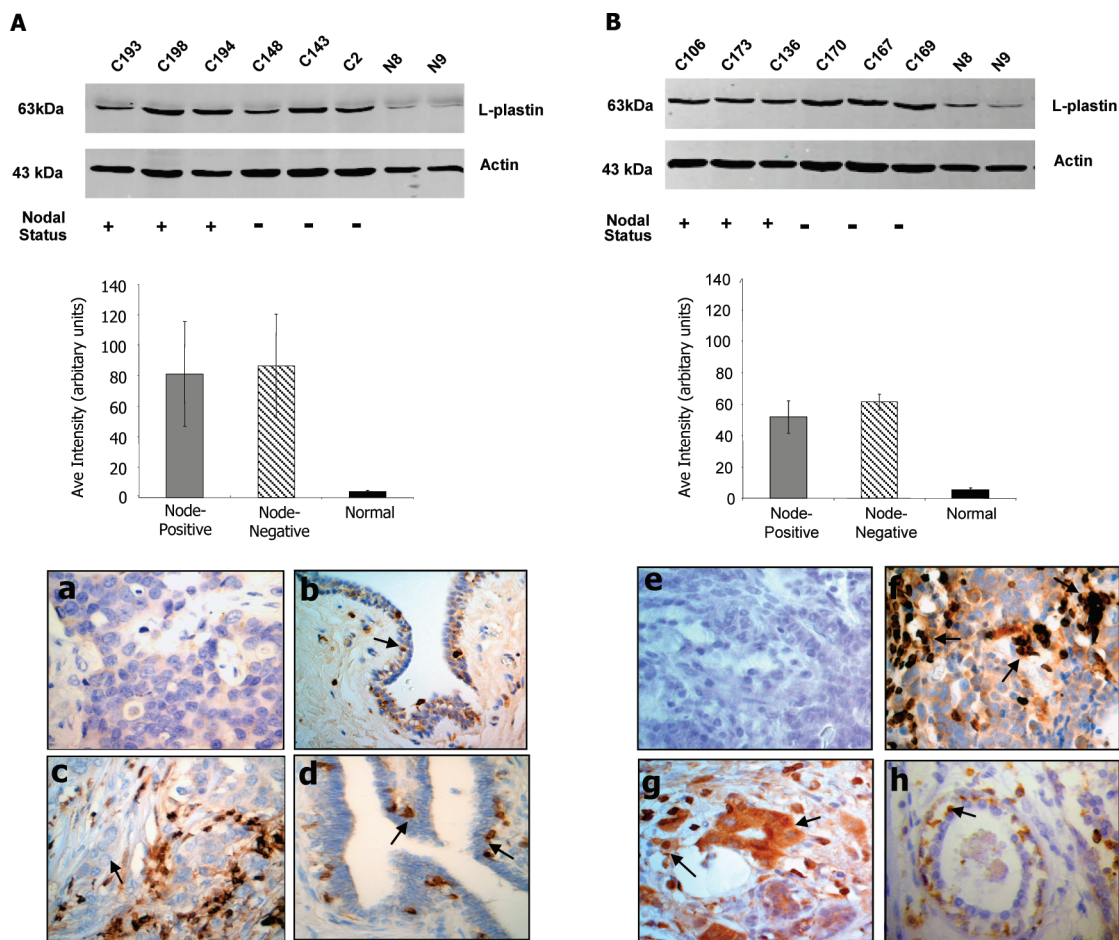
Figure 5. L-Plastin protein expression in breast carcinomas. (A) Expression of L-plastin in ER+/PgR+ receptor carcinomas. Upper panel, Western blot analysis; middle panel, densitometry (*n* = 3 individual cases per group); lower panel, IHC (original magnification ×40) (a) negative control, (b) normal tissue (arrows indicate myoepithelial cells), (c) malignant cells (arrows indicate nonstained malignant cells), and (d) malignant tissue (arrows indicate positive staining of myoepithelial cells). (B) Expression of L-plastin in ER−/PgR− receptor carcinomas. Upper panel, Western blot analysis; middle panel, densitometry; lower panel, IHC, original magnification ×40, (e) negative control, (f) lymphocytes (indicated by arrows), (g) cytoplasmic and nuclear staining, (h) myoepithelial staining.

A very interesting result is the identification of the homeodomain protein nanog. This protein is expressed in embryonic stem cells and was recently shown to be necessary and sufficient for their pluripotency.[27]

2. Membrane Protein Profiles of Tumors That Have Metastasized to the Lymph Nodes versus Tumors That Have Not Metastasized to the Lymph Nodes: Methodological Potential for Detection of Biomarkers of Early Metastasis. A pool of the membrane proteomes of 3 cancers that have metastasised to the lymph nodes were compared to a pool of the membrane proteomes of 3 nonmetastatic cancers. Membrane proteins were used to increase the possibility of discovering proteins that are involved in signal transduction/adhesion/migration, that would be easily accessible as biomarkers and to decrease the noise from the millions of distinct proteins that are in a cell at anyone time. Table 2 lists 29 proteins detected only in the membrane fraction of the nonmetastatic cancer sample pool. L-Plastin, although not an integral membrane protein, is again identified as a protein specifically overexpressed in the nonmetastatic tumor sample. This is not consistent with validation experiments where L-plastin is seen in metatstatic tumors by Western blot and IHC. However, this may be an interesting observation suggesting that L-plastin has a tighter association with the membrane in

nonmetastatic tumors, indicating a spatially regulated role in metastasis. Table 3 lists 22 proteins that were detected in the metastatic tumor sample but not in the nonmetastatic sample. An intriguing example of a candidate marker for invasiveness is lactadherin, a protein that promotes cancer cell adhesion via integrins.[28] We detected this protein with 5 peptides in the membrane fraction of the metastatic cancer sample pool but did not detect it in the nonmetastatic cancer sample pool. The result suggests that lactadherin is much more abundant in metastatic tumors and could potentially be used as a marker for invasiveness.

3. Distribution of the Identified Candidate Markers Across Molecular Function and Biological Process Classes: A Comparison with the Human Proteome. Figure 3 presents a statistical assessment of the obtained results based on a comparison with the Human proteome. Gene ontology (GO) was used to classify the identified candidate marker proteins by molecular function and biological process. The top panel shows the enriched biological process categories. Most of the strongly enriched terms associate with regulation of the cytoskeleton, motility, and organelle biogenesis and localization. The bottom panel shows the results of a similar analysis but based on the GO term molecular function. Strongly enriched functions are Actin binding, peroxiredoxin activity, calmodulin
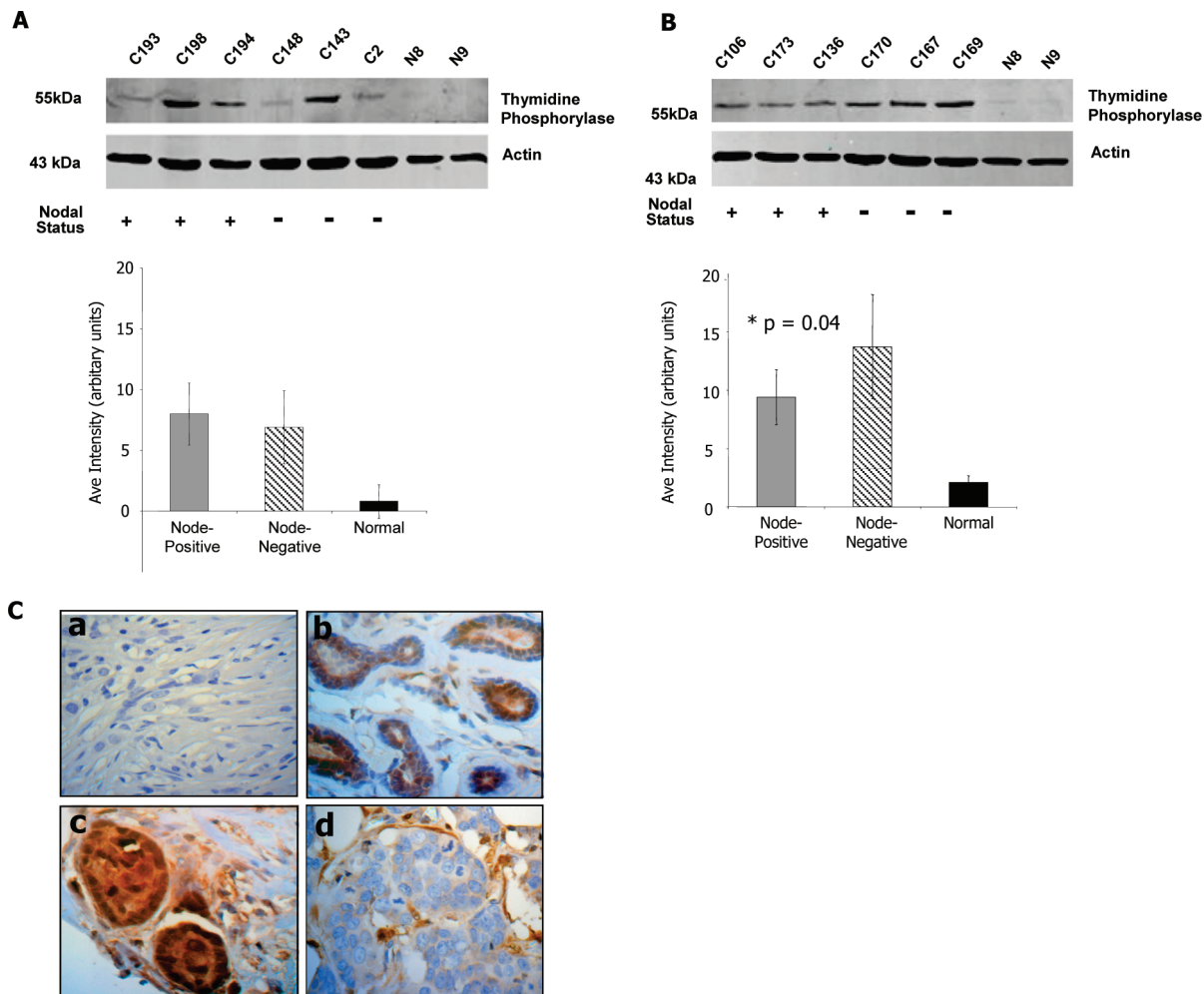
Figure 6. Thymidine phopsphorylase protein expression in breast carcinomas. (A) Expression of thymidine phosphorylase in ER+/PgR+ receptor carcinomas. Upper panel, Western blot analysis; lower panel, densiometric analysis ($n = 3$ individual cases per group). (B) Expression of thymidine phopsphorylase in ER−/PgR− receptor carcinomas. Upper panel, Western blot analysis; lower panel, densiometric analysis ($n = 3$ individual cases per group). (C) IHC showing cellular distribution and intensity of thymidine phosphorylase in tumors of differing grade, original magnification ×40, (a) negative control, (b) normal tissue showing expression in epithelial and macrophage cells, (c) a representative grade 2 tumor (IDC +/−DCIS ER+/PgR+) showing expression in malignant cells, (d) a representative grade 3 tumor (IDC +/−DCIS ER+/PgR+) showing no expression in the malignant cells and some lymphocytic expression.

binding, and intramolecular oxidoreductase activity. A table with the complete set of results from the GO analysis is available in the Supporting Information.

In summary, two groups of proteins are strongly over presented in the breast cancer data set, the proteins involved in response to stress and the proteins involved in cellular motility and regulation of the cytoskeleton. It is worth noting that these two groups exemplify functions that are critical for the survival of the tumor cells and for their invasiveness and metastasis.

Validation Experiments. To check the validity of the results obtained by mass spectrometry, we conducted Western blotting and immunohistochemical staining experiments using antibodies specific for several of the proteins detected only in the tumors but not in healthy or peripheral tissue samples. L-Plastin, thymidine phosphorylase (TP), STAT1, CLC1, and Nanog were chosen for analysis due to relevance considerations and availability of good antibodies.

STAT1 and Nanog showed diverse patterns of expression with no apparent correlation with nodal status or histological grade (Figure 4). Shen-chen et al.[26] also found no correlation
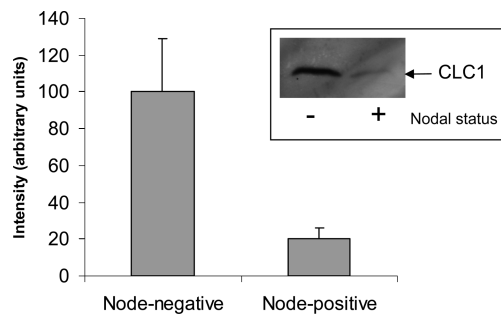

Figure 7. CLC1 analysis by Western blotting in membrane fractions of metastatic and nonmetastatic cancers. Membrane proteins were enriched as described in Experimental Procedures. CLC1 was detected with anti-CLC1 mouse monoclonal primary antibody and fluorescent antimouse secondary antibody. (Bar graph) Quantitative analysis of CLC1 abundance in node-positive membrane protein isolates and node-negative membrane protein isolates. Fluorescence intensity was determined using the Odyssey software. Values are mean of 4 individual node-positive cancers and 4 individual node-negative cancer samples. (Inset) A representative Western blot analysis of pooled node-positive ($n = 3$) and node-negative ($n = 3$) membrane protein samples.

of STAT1 expression with age, receptor status, grade, lymph node metastasis, and overall survival rate. Nanog is a transcription factor that plays a key role in self-renewal and maintenance of pluripotency in human embryonic stem cells. Links to breast cancer include the expression of NANOGP8 in cancer cell lines and cancer tissues[31] and in breast carcinomas.

L-Plastin was significantly more abundant in all cancers tested (ILC and IDC II, data not shown) compared to normal samples. However, no significant difference between metastatic and nonmetastatic tissue was seen in Western blots (Figure 5A). L-Plastin is a protein specific for blood cells but is also known to be expressed in some breast cancers cell lines. It has been detected with antibodies in many solid tumors including breast tumors.[13] Although it is a protein already implicated in breast cancer, it could be a useful validation experiment to study its expression in our panel of breast cancer specimens. In particular, we were interested to determine to what extent detection of L-plastin in cancer samples was due to leukocyte infiltration rather than endogenous expression by the malignant cells. Immunohistochemitry experiments (Figure 5B) revealed that the majority of L-plastin expression was seen in infiltrating leukocytes and benign myoepithelia, which highlights the problem of interpretation of results from hetreogenous tissue. However, we found that ER/PgR-negative, node-negative IDC breast tumors did express L-plastin in the malignant cells. This result is similar to the data reported by Lapillone et al.[13] with the exception that we do not detect L-plastin in normal epithelial tissue by immunohystochemical staining.

Thymidine phosphorylase was specifically detected in all types of breast cancer tissue (ILC and IDC II, data not shown) but not in normal tissue. The protein showed a trend (although not highly significant with this sample size) of increased expression in node-negative ER−/PgR− tumors compared to node-positive tumors with the same receptor status (Figure 6A). Immunohistochemistry experiments (Figure 6 B) revealed that the protein was expressed in the nucleus and the cytoplasm of the malignant cells. There was also some stromal staining, although this appeared to be fibrous and not cellular. Macrophages also expressed thymidine phosphorylase. This protein showed an interesting pattern of expression in malignant cells that correlated with the grade of the tumor. Malignant cells in lower grade tumors have abundant expression compared to the malignant cells in the higher grade tumors. This may allude to the potential of using thymidine phosphorylase as a prognostic indicator. This is supported by the work of Ioachim et al., who found that high levels of thymidine phosphorylase were significantly associated with a favorable outcome.[9] The validation data for these two proteins illustrate some of the advantages and pitfalls of proteomic approaches to discovery. The heterogeneity of the tissue can yield results that are irrelevant to the malignant cells. However, with thorough post-MS validation, interesting cell-specific patterns can be discovered that correlate with related pathologies of breast cancer.

CLC1 was not detected by Western blotting in total tissue lysates. However, as shown on Figure 7, CLC1 was consistently found to be more abundant in the membrane fractions isolated from nonmetastatic tumors compared to the membrane fractions from metastatic tumors. This supports the data obtained by mass spectrometry (Table 2) and suggests that early metastasis may be associated with decreased expression of this protein in the membranes of, as yet unspecified, cell types in the heterogeneous tumor tissue.

Notes on Methodology. Proteomic technologies, including bioinformatic analysis tools, have progressed to a level to allow their application to solve clinical problems. It is now possible to identify pathological components of a disease by large-scale analysis of expressed proteins. Advances in protein biochemistry as well as robust established techniques allow validation, monitoring, and functional analysis of discovered protein candidates. Despite this, no protein identified by proteomics has as yet made it to the clinic as a viable breast cancer biomarker, suggesting that better fusion of the postgenomic research methodology with clinical practice is needed. One of the strengths of our approach is the synergy of clinical and proteomic studies from the outset. The use of rigorously collected human samples together with prior and follow-up pathological/clinical and life-style data should reduce variability due to sample origin, collection, and handling, and should bring the discoveries obtained by proteomics a step closer to the establishment of novel markers and drug target candidates.

A large variety of experimental strategies and approaches based on mass spectrometry are now available for the needs of cancer proteomics, each with its own strength and weaknesses (reviewed in refs 1 and 20). In this study, we have chosen an approach that combines 1D-SDS-PAGE and nanoscale LC/MS/MS. 1D-SDS-PAGE is very robust and reproducible technique and is by far the highest resolution protein separation method. Nano-LC/MS has matured and is one of the most powerful platforms for ultra sensitive analysis of peptides. In this study, 1D-SDS-PAGE served as initial top-down fractionation step prior to the LC/MS/MS analysis. The method is equally appropriate for the separation of soluble and membrane proteins, for highly acidic and basic proteins. It provides information for the size of the separated proteins, which is important in the subsequent analysis. In previous years, 1D-SDS-PAGE was regarded as inferior to 2-DE, because of the need to use peptide mass fingerprinting for the subsequent identification of proteins. Peptide mass fingerprinting (PMF) requires highly purified sample; hence, 1D-SDS-PAGE where any given band can contain as many as 100 detectable proteins was not the method of choice. 2-DE in combination with MALDI-TOF and fingerprinting is still the predominant technology in many proteomics laboratories. Its inherent lack of throughput is ameliorated to some extent by the use of robotic spot pickers and automatic digestion and spotting stations, but this cannot circumvent the low sensitivity and moreover the lower reliability of protein identification associated with PMF. Several alternatives to 2-DE/MALDI have been developed to solve these problems. An example is the MuDPIT method which does not include any separation of intact proteins.[30] Although very powerful, this method has its own shortcomings; since protein inference is based on peptides detected and identified by MS/MS analysis of protease digests of unseparated protein samples, all information about isoforms and splice variants is lost in this analysis. There are also technical problems accompanying the attempt to analyze many millions of peptides in a single LC run. Ionization suppression, overloading of the system, and poor reproducibility are some of the known problems.

In this study, 1D-SDS-PAGE was used to generate size-resolved fractions of the proteomes of cancer, normal and peripheral to cancer tissue samples. The proteins in these fractions were digested and the generated peptides were analyzed by a fully automatic system consisting of an autosampler, nanoscale separation module, and a fast-scanning, high-

capacity ion trap mass spectrometer. The tandem mass spectra generated from each fraction were converted to mzXML files[21] and uploaded automatically into a data analysis pipeline based on open source tools such as CPAS[4,22] and the Transproteomic Pipeline.[11] This experimental setup affords the identification of as many as 2000 proteins in a single day. Several hundred of these proteins are detected with multiple peptides and are quantified automatically by the pipeline using the spectral count method.[15] The outcome of such an analysis is a set of differential displays which reveal the patterns that distinguish healthy from diseased specimens and can be used to establish multimodal biomarkers and to discover drug target candidates. Unlike other methods that have been applied to breast tumor samples, such as SELDI for example,[17] MS/MS delivers true protein identities, and more importantly, the relative abundance of these proteins can be estimated. Prefractionation by SDS-PAGE is a very useful and reproducible way to decrease sample complexity. As a top-down approach, it has the advantage of preserving the information about protein isoforms such as splice variants, unlike MuDPIT and other shot-gun methods. It is worth noting also that in all analyzed samples we detected large amounts of several serum proteins. Most notably, these were albumin, transferin, and various immunoglobulins. This shows that the efficiency and potential of the SDS-PAGE−LC/MS approach could be increased dramatically by using an additional prefractionation step to deplete these high-abundance proteins from the samples. We expect that such depletion will increase the number of reliably quantifiable proteins to 1000 and more.

An important feature of the approach implemented in this study is the use of label-free quantitation. Although this method is not as precise as stable isotope based methods such as ICAT, ITRAQ, and IPCL,[8,23,3] it is good enough to detect differences in abundance as small as 30%.[19] This is sufficient for the needs of a typical biomarker discovery project. In fact, it might be that it is sufficient to only stratify the identified proteins into relatively broad abundance groups and then select for validation protein hits that are found in some of the high-abundance group in tumors but are not detected at all or detected with 1–2 spectra in the control sample. Label-free quantitation has an additional advantage over isotope-based techniques. It allows an unlimited number of experiments to be cross compared even when some of these experiments have been performed long periods of time apart. This is essential as our strategy is to continue to acquire data from tumors and control tissue over a long period of time. At present, we have collected over 200 different tumor specimens and matching periphery tissue samples. The tumors are grouped according to their phenotype, grade, size, ER/PgR status, HER2 expression, and nodal status. We also have samples from patients that have undergone primary chemotherapy showing differing responses, for analysis of proteins involved in resistance to chemotherapy. All these samples will give rise to a data set of many millions of MS/MS spectra, and the use of label-free quantitation will allow us to use a simple unbiased approach to compare the abundance of all identified proteins across all samples and proteome fractions.

Concluding Remarks

We have combined robust and mature analytical techniques and advanced bioinformatics tools into an integrated research platform for identification and validation of breast cancer biomarkers using freshly collected, fully characterized human breast tissue. With this approach, we have identified, quantified, and validated proteins from samples derived from different types of tumor, tissue peripheral to the tumor, and normal breast tissues. So far we have collected over 500 000 MS/MS spectra from size-resolved proteomes from selected phenotypically matched tissue samples which led to identification of established and novel breast cancer protein biomarker candidates, demonstrating the utility and effectiveness of the approach. Selected proteins were studied by Western blotting and immunohistochemistry techniques to validate the results obtained by mass spectrometry. We are currently using this platform to analyze a cohort of over 200 tumors and matching periphery tissue samples for further validation and discovery and to establish a growing tumor proteome database.

Supporting Information Available: Tables listing the identified proteins with corresponding high-score peptides from 1 cancer and matched periphery, the complete set of results from GO analysis, and the complete output of the GOLEM analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207.

(2) Boyle, E. I.; Weng, S.; Gollub, J.; Jin, H.; Botstein, D.; Cherry, J. M.; Sherlock, G. GO TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics 2004* **2004**, *20* (18), 3710–3715.

(3) Chong, P. K.; Gan, C. S.; Pham, T. K.; Wright, P. C. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.* **2006**, *5*, 1232–1240.

(4) Cottingham, K. CPAS: a proteomics data management system for the masses. *J. Proteome Res.* **2006**, *5*, 14.

(5) Cove, D. H.; Woods, K. L.; Smith, S. C.; Burnett, D.; Leonard, J.; Grieve, R. J.; Howell, A. Tumor markers in breast cancer. *Br. J. Cancer* **1979**, *40*, 710–871.

(6) Ezeh, U. I.; Turek, P. J.; Reijo, R. A.; Amander, T. C. Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer* **2005**, *104*, 2255–2265.

(7) Foran, E.; McWilliam, P.; Kelleher, D.; Croke, D. T.; Long, A. The leukocyte protein L-plastin induces proliferation, invasion and loss of E-cadherin expression in colon cancer cells. *Int. J. Cancer* **2006**, *118*, 2098–2104.

(8) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.

(9) Ioachim, E. Thymidine phosphorylation expression in breast cancer: the prognostic significance and its association with other angiogenesis related proteins and extracellular matrix componenets. *Histol. Histopathol.* **2008**, *23*, 187–196.

(10) Jacobs, J. M.; Adkins, J. N.; Qian, W. J.; Liu, T.; Shen, Y.; Camp, D. G., II; Smith, R. D. Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* **2005**, *4*, 1073–1085.

(11) Keller, A.; Eng, J.; Zhang, N.; Li, X. J.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **2005**, *1*, 0017.

(12) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685.

(13) Lapillonne, A.; Coue, O.; Friederich, E.; Nicolas, A.; Del Maestro, L.; Louvard, D.; Robine, S.; Sastre-Garau, X. Expression patterns

of L-plastin isoform in normal and carcinomatous breast tissues. *Anticancer Res.* **2000**, *20*, 3177–3182.

(14) Lin, C. S.; Lau, A.; Yeh, C. C.; Chang, C. H.; Lue, T. F. Upregulation of L-plastin gene by testosterone in breast and prostate cancer cells: identification of three cooperative androgen receptor-binding sequences. *DNA Cell. Biol.* **2000**, *19*, 1–7.

(15) Liu, H.; Sadygov, R. G.; Yates, J. R., III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193–4201.

(16) Maclean, B.; Eng, J. K.; Beavis, R. C.; McIntosh, M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* **2006**, *22* (22), 2830–2832.

(17) Mathelin, C.; Cromer, A.; Wendling, C.; Tomasetto, C.; Rio, M. C. Serum biomarkers for detection of breast cancers: a prospective study. *Breast Cancer Res. Treat.* **2006**, *96*, 83–90.

(18) Metodiev, M. V.; Timanova, A.; Stone, D. E. Differential phosphoproteome profiling by affinity capture and tandem matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* **2004**, *4*, 1433–1438.

(19) old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4*, 1487–1502.

(20) Patterson, S. D.; Aebersold, R. H. Proteomics: the first decade and beyond. *Nat. Genet.* **2003**, *33* (Suppl.), 311–323.

(21) Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* **2004**, *22*, 1459–1466.

(22) Rauch, A.; Bellew, M.; Eng, J.; Fitzgibbon, M.; Holzman, T.; Hussey, P.; Igra, M.; Maclean, B.; Lin, C. W.; Detter, A.; Fang, R.; Faca, V.; Gafken, P.; Zhang, H.; Whiteaker, J.; States, D.; Hanash, S.; Paulovich, A.; McIntosh, M. W. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J. Proteome Res.* **2006**, *5*, 112–121.

(23) Schmidt, A.; Kellermann, J.; Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5*, 4–15.

(24) Sealfon, R. S.; Hibbs, M. A.; Huttenhower, C.; Myers, C. L.; Troyanskaya, O. G. GOLEM: an interactive graph-based gene-ontology navigation and analysis tool. *BMC Bioinf.* **2006**, *7*, 443.

(25) Seo, J.; Shneiderman, B. Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework. *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 311–322.

(26) Sheen-Cheen, S. M.; Huang, C. C.; Tang, R. P.; Yang, C. H.; Chou, F. F.; Eng, H. L. Signal transducer and activator of transcription 1in breast cancer: analysis with tissue microarray. *Anticancer* **2007**, *27*, 2481–2486.

(27) Silva, J.; Chambers, I.; Pollard, S.; Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **2006**, *441*, 997–1001.

(28) Taylor, M. R.; Couto, J. R.; Scallan, C. D.; Ceriani, R. L.; Peterson, J. A. Lactadherin (formerly BA46), a membrane-associated glycoprotein expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp (RGD)-dependent cell adhesion. *DNA Cell. Biol.* **1997**, *16*, 861–869.

(29) Toi, M.; Ueno, T.; Matsumoto, H.; Saji, H.; Funata, N.; Koike, M.; Tominaga, T. Significance of thymidine phosphorylase as a marker of protumor monocytes in breast cancer. *Clin. Cancer Res.* **1999**, *5*, 1131–1137.

(30) Wolters, D. A.; Washburn, M. P.; Yates, J. R., III. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683–5690.

(31) Zhang, J.; Wang, X.; Li, M.; Han, J.; Chen, B.; Wang, B.; Dai, J. NANOGP8 is a retrogene expressed in cancers. *FEBS* **2006**, *273*, 1723–1730.

PR7007829