Somiari, R.I. et al. High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. Proteomics 3, 1863-1873

ARTICLE *in* PROTEOMICS · OCTOBER 2003

Impact Factor: 3.81 · DOI: 10.1002/pmic.200300560 · Source: PubMed

Richard I. Somiari[1]
Anthony Sullivan[1]
Stephen Russell[1]
Stella Somiari[1]
Hai Hu[1]
Rick Jordan[1]
Alisha George[1]
Richard Katenhusen[1]
Alicja Buchowiecka[1]
Cletus Arciero[2]
Henry Brzeski[1, 2]
Jeff Hooke[2]
Craig Shriver[2]

[1]Clinical Breast Care Project, Windber Research Institute, Windber PA, USA
[2]General Surgery Services, Walter Reed Army Medical Center, Washington DC, USA

High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast

Large-scale proteomics will play a critical role in the rapid display, identification and validation of new protein targets, and elucidation of the underlying molecular events that are associated with disease development, progression and severity. However, because the proteome of most organisms are significantly more complex than the genome, the comprehensive analysis of protein expression changes will require an analytical effort beyond the capacity of standard laboratory equipment. We describe the first high-throughput proteomic analysis of human breast infiltrating ductal carcinoma (IDCA) using OCT (optimal cutting temperature) embedded biopsies, two-dimensional difference gel electrophoresis (2-D DIGE) technology and a fully automated spot handling workstation. Total proteins from four breast IDCAs (Stage I, IIA, IIB and IIIA) were individually compared to protein from non-neoplastic tissue obtained from a female donor with no personal or family history of breast cancer. We detected differences in protein abundance that ranged from 14.8% in stage I IDCA *versus* normal, to 30.6% in stage IIB IDCA *versus* normal. A total of 524 proteins that showed $\geq$ three-fold difference in abundance between IDCA and normal tissue were picked, processed and identified by mass spectrometry. Out of the proteins picked, $\sim$80% were unambiguously assigned identities by matrix-assisted laser desorbtion/ionization-time of flight mass spectrometry or liquid chromatography-tandem mass spectrometry in the first pass. Bioinformatics tools were also used to mine databases to determine if the identified proteins are involved in important pathways and/or interact with other proteins. Gelsolin, vinculin, lumican, alpha-1-antitrypsin, heat shock protein-60, cytokeratin-18, transferrin, enolase-1 and β-actin, showed differential abundance between IDCA and normal tissue, but the trend was not consistent in all samples. Out of the proteins with database hits, only heat shock protein-70 (more abundant) and peroxiredoxin-2 (less abundant) displayed the same trend in all the IDCAs examined. This preliminary study demonstrates quantitative and qualitative differences in protein abundance between breast IDCAs and reveals 2-D DIGE portraits that may be a reflection of the histological and pathological status of breast IDCA.

Keywords: High-throughput proteomics / Infiltrating ductal carcinoma / Two-dimensional difference gel electrophoresis
PRO 0560

1 Introduction

Proteomics includes the large-scale global analysis of proteins expressed by the genome of an organism. The typical aim being to look for quantitative changes that occur as a function of disease, treatment or environment. Although genomic and mRNA profiling technologies will continue to provide valuable insights into the biology of cancers, there are limits to the type and amount of information that can be generated at the DNA and mRNA levels [1] and transcriptional mechanisms do not always mirror translation [2, 3]. Large-scale proteomics will play a critical role in the rapid identification and validation of new protein targets [4], and elucidation of the underlying molecular events associated with disease development, progression and severity [5–7]. However, because of the complexity of the proteome of most organisms, the systematic and complete cataloging of protein expression changes will require an analytical effort beyond the capacity of standard laboratory equipment and throughput.

Correspondence: Richard I. Somiari, PhD., Clinical Breast Care Project, Windber Research Institute, 600 Somerset Avenue, Windber PA 15963, USA
E-mail: r.somiari@wriwindber.org
Fax: +1-814-467-6334

Abbreviations: Cy, cyanine; **DIA**, difference in-gel analysis; **DIGE**, difference gel electrophoresis; **ER**, estrogen receptor; **HER-2**, human epidermal growth factor receptor-2; **IDCA**, infiltrating ductal carcinoma; **LC**, low-in-cancer; **LN**, low-in-normal; **OCT**, optimal cutting temperature; **PR**, progesterone receptor

Proteomics strategies have been used to identify cancer-specific protein markers that could provide the basis for the development of new methods for diagnosis, treatment and early disease detection [6, 8–16]. It is apparent from the number of reports in literature that two-dimensional electrophoresis (2-DE) is still the benchmark for large-scale separation of complex protein mixtures. 2-DE currently offers the best resolution possible because it separates proteins according to two independent physicochemical parameters: isoelectric point and size [4, 11, 17, 18] and the gel serves as an efficient fraction collector that protects the protein molecules until picked [4]. The application of 2-DE in proteomics has greatly expanded beyond the original concept of simply providing a complete picture of the proteins expressed in individual samples. Recent developments have led to the availability of a spectrum of 2-DE specific reagents, precast gels, immobilized pH gradient strips, multiple detection and identification techniques and bioinformatics tools for automated and quantitative image analysis [1, 4, 19, 20]. Many modifications of 2-DE have significantly improved reproducibility, sensitivity and resolution [21–24]; it is still difficult to fully duplicate the pattern of protein expression using conventional 2-DE methods [11, 25].

The recently introduced variety of 2-DE, termed difference gel electrophoresis (DIGE) promises to significantly improve the speed, reproducibility, sensitivity of 2-DE-based proteomics and ease the drudgery associated with comparing multiple gels, homologous protein spots and comparative experiments [11, 26, 27]. The DIGE concept involves the covalent labeling of each pool of protein extract with different fluorescent dyes such as cyanine dyes [11, 25–28] or Alexa dyes [25]. Equal concentrations of the differentially labeled protein samples are mixed and coseparated during the same 2-DE process. The 2-DE gel pattern is then visualized by imaging of the gel with a fluorescence scanner by sequential excitation of the fluorescence dyes used. The protein : dye ratio is deliberately kept high (>95%) so that only the proteins containing a single dye molecule are visualized on the gel [27]. The charge and mass of the fluorescence dyes used in 2-D DIGE are carefully matched to reduce dye-induced shift of proteins during 2-DE [27]. A comparison of the images generated by scanning of the DIGE gel at two wavelengths allows the quantitation of each spot [11, 25–27] and the use of a third dye permits the comparison of multiple gels [28], thereby reducing the reproducibility problem. Quantitative analysis of protein profiles is reported to be fast and accurate because it is based on the relative fluorescence intensities captured from a single gel. 2-D DIGE has been used successfully for proteomic analysis of proteins from *Escherichia coli* cell extract, colorectal adenocarcinoma cell lines, mouse liver homogenates, esophageal tissue sections, breast ductal carcinoma *in situ*, and breast cancer cell systems overexpressing ErbB-2 [11, 12, 25–28]. The accompanying DeCyder software (Amersham Biosciences, Piscataway, NJ, USA) clearly distinguishes protein expression differences that were statistically significant and correlate with data obtained by Western blotting [11, 28]. The comparison of the spot intensities using the difference in-gel analysis module of the DeCyder software is relatively more objective than the conventional approaches involving manual adjustment of the brightness/contrast of two gel images [11].

The recent proteomic analysis of human breast ductal carcinoma *in situ* (DCIS) revealed novel protein expression patterns or modification trends that were distinct from the results of nucleic acid based approaches [12], suggesting that the proteomic portrait of human breast IDCAs will likely be different from the results obtained by mRNA and DNA analysis. We are not aware of any paper describing the use of 2-D DIGE technologies for proteomic analysis of human breast infiltrating ductal carcinoma (IDCA). We are actively studying breast infiltrating ductal carcinomas (IDCAs) because this disease imposes significant cancer burden on women worldwide. Elucidation of the proteins that are differentially expressed in IDCAs could lead to a better understanding of the complex molecular events that are associated with breast cancer progression and cellular signaling as well as potentially providing the basis for the development of new methods to manage breast IDCA. In this study, protein extracts from test (IDCA, $n = 4$) and reference (reductive mammoplasty, $n = 1$) samples were labeled with fluorescent dyes, mixed and subjected to 2-D DIGE. The fluorescence signals generated after 2-D DIGE were analyzed and the differentially expressed proteins picked and processed with a fully automated spot handling workstation and identified by mass spectrometry. Subsets of the identified proteins were then used as queries to search multiple databases of interacting proteins using proprietary and commercially available bioinformatics tools. This work represents the first high-throughput proteomic analysis of human breast IDCA using the 2-D DIGE technologies and a fully automated spot handling workstation.

2 Materials and methods

2.1 Sample processing

Breast tumors classified as IDCAs were obtained from the Institutional Review Board (IRB) approved tissue bank jointly setup by the Clinical Breast Care Project (CBCP) of Walter Reed Army Medical Center, Washington DC,

USA; Joyce Murtha Breast Care Center, Windber Medical Center, Windber PA, USA and Windber Research Institute, Windber PA, USA. Specimens were obtained from fully informed and consenting donors, and only used for research projects approved by the IRB of all the institutions. Board certified clinical oncologists and pathologists carried out clinical consultation and histological analysis of biopsies and all the samples were made anonymous before transport to the Windber Research Institute tissue banking facility. Each specimen was collected using stringent and uniform standard operating protocols developed by the CBCP. Typically, biopsies were placed in sterile containers within 3 min of a medically indicated surgical procedure and held on ice until embedded in OCT (optimal cuttung temperature) medium. The time from the removal of lesions by surgery to embedding in OCT did not exceed 1 h. For this study, we used four poorly differentiated IDCAs with known histological stage, estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER-2/neu Erbb2) status.

For proteomic analysis, multiple 8–10 μm sections were shaved off each OCT embedded tumor in a cryostat. The first, middle and last sections were stained with hematoxylin and eosin (H&E) and examined by a pathologist to ensure that at least 70% of the tissue sections used for the proteomic study contained tumor cells. At least eight shavings sandwiched by sections verified histologically were pooled, suspended and homogenized in 100 μL of 2-DE lysis buffer (7 M urea, 2 M thiourea, 30 mM Tris, 5 mM MgAc, 4% CHAPS, 1% NP-40), incubated at room temperature (23 ± 1°C) for 30 min with vortexing at ~10 min intervals. The homogenate was centrifuged at 14 000 × *g* for 10 min and the clear supernatant transferred to a clean, and sterile microcentrifuge tube and stored at −80°C until needed.

2.2 Sample preparation for 2-D DIGE

Total protein concentration was determined by the micro Bradford assay (Pierce, Rockford, IL, USA), and similar protein concentrations were used for two dye 2-D DIGE analysis. Protein from four breast IDCAs (WRI-329, WRI-369, WRI-393, WRI-457) were individually compared to normal breast tissue obtained after reductive mammoplasty. Approximately 50 μg of each test protein (IDCA) suspended in lysis buffer was labeled with 200 pmol of cyanine 3 (Cy3), and 50 μg of reference protein (reductive mammoplasty) was labeled with 200 pmol of Cy5 according to the manufacturer's protocol (Amersham Biosciences). After labeling, 125 μg of test and 125 μg of reference protein were added (unlabeled) to the respective pools, to bring the final concentration of each pool to 175 μg. The additional protein was added to increase the total protein concentration, eliminate the need to perform separate preparative 2-D DIGE in parallel, and enable downstream processing and analysis of differentially expressed proteins by peptide mass fingerprinting. The total amount of protein used for each 2-D DIGE experiment was 350 μg (test, 175 μg; reference, 175 μg).

2.3 2-D DIGE analysis

The differentially labeled specimens (test and reference) were combined and mixed with 450 μL of rehydration buffer (2 M thiourea, 7 M urea, 4% CHAPS, 1% NP-40, 10% isopropanol, 5% glycerol, 0.5% 3–10 IPG buffer and 20 mM DTT), before being applied to a 24 cm IPGphor strip (pH 3–10) in a strip holder for rehydration at 30 V for 12 h. The first dimension separation was carried out using an IPGphor isoelectric focusing system (Amersham Biosciences). The total focusing time was 65 500 Vh and the strips were equilibrated in two steps: (i) 15 min in a solution containing 6 M urea, 50 mM Tris (pH 8.8), 30% glycerol, 2% SDS and 0.5% DTT; and (ii) 15 min in a similar solution containing 4.5% iodoacetamide rather than DTT. After equilibration, second dimension electrophoresis was performed in an ETTAN Dalt 6 (Amersham Biosciences) at 160 W for 6 h, using 12% isocratic polyacrylamide gels previously cast in 20 × 24 mm glass plates, of which one inner side was precoated with bind-silane. After 2-DE, gels were scanned on the Typhoon 9400 variable mode imager (Amersham Biosciences) using excitation/emission wavelengths specific for Cy3 (532 nm/580 nm) and Cy5 (633 nm/670 nm). Differentially expressed proteins were identified with the DIA module of DeCyder. Images were normalized, statistically analyzed and differentially expressed proteins identified and quantified with DeCyder-DIA (Amersham Biosciences) using the Cy dye generated images. The normalized spot volumes generated in DeCyder-DIA and showing ≥threefold difference between the test and reference samples were summed up and the overall difference in spot volume between the IDCAs tested for statistical significance by *t*-test using Excel™ (Microsoft). After DIA, each gel was stained with SYPRO Ruby (Molecular Probes, Eugene, OR, USA) and the image captured using an excitation wavelength of 457 nm and emission wavelength of 610 nm. A pick list was generated from the SYPRO Ruby stained gels using DeCyder-DIA.

2.4 Automated protein spot picking and in-gel digestion

The spot picking, destaining, digestion, extraction, MALDI sample preparation and spotting on MALDI target slides were carried out robotically in an enclosed fully

automated spot handling workstation (ETTAN ProSpot; Amersham Biosciences). Briefly, each picked gel plug was rinsed by three wash cycles in which 150 μL 50% methanol/50% water, containing 15 mM ammonium bicarbonate was applied, left to stand for 15 min, and then removed. The plugs were dehydrated in 75% acetonitrile for 20 min, the solution removed, and the trays placed into a heated, air circulating drying module and dried for 10 min. To the dried plugs was added 7 μL of digestion solution (containing 40 ng/μL sequencing grade trypsin in 15 mM ammonium bicarbonate), and the plate was robotically transferred to the digestion station where it was covered and maintained at 37°C for 1 h. Eighty microliters of extraction solution (50% acetonitrile containing 0.1% TFA) were then added to each well and left to stand for 20 min. The extract solutions were transferred to a clean 96-well plate, and the extraction repeated. The plate containing the combined extracts was then evaporated to dryness. MALDI samples were prepared by reconstituting the dried peptides in 5 μL of 50% acetonitrile containing 0.5% TFA. An aliquot of 0.3 μL was then applied to the clean MALDI target slide surface and allowed to dry. The spots were redissolved by the addition of 0.3 μL of matrix solution (α-cyano-4-hydroxycinnamic acid saturated solution in 50% acetonitrile containing 0.5% TFA) and finally allowed to dry before MALDI-MS. The plate containing the remaining peptides was stored at −20°C until LC-MS/MS analyses could be performed on the samples that were not identified by MALDI-MS.

2.5 Mass spectrometric identification of proteins

MALDI peptide mass fingerprinting was carried out on an ETTAN PRO MALDI-TOF mass spectrometer operating in reflectron mode (Amersham Biosciences). Internal calibration was performed using the trypsin autodigestion peaks at *m/z* 842.509 and 2211.104. Each spectrum corresponded to the sum of 255 acquisitions, each of 8 laser pulses, in which the threshold signal-to-noise exceeded a set value. Protein identification by peptide mass fingerprinting was performed using the MASCOT search engine (Matrix Science, London, UK) and the National Center for Biotechnology Information protein database (NCBInr). Proteins that were not resolved by MALDI-MS were subjected to LC-MS/MS using a capillary HPLC system and a Q-TOF API US mass spectrometer (Waters, Milford, MA, USA), using the following solvent systems: A (95% acetonitrile, 5% water, 0.1% formic acid) and B (5% acetonitrile, 95% water, 0.1% formic acid). Peptide mixtures were reconstituted in 10 μL of 95% water and 5% acetonitrile containing 0.1% formic acid. Five microliters of this solution were injected by microliter pickup into a 20 μL injection loop and transferred onto a 5 mm C18 trap column. After 3 min, the flow to the trap column was reversed, and a gradient (95% A to 30% A over 40 min) was applied to separate and elute the peptides onto a PepMap C18 capillary column (Dionex, Sunnyvale, CA, USA) at a flow rate of ~250 nL/min. The mobile phase composition was set at 85% B and held for 5 min to elute all remaining peptides, then adjusted back to initial conditions (95% A, 5% B) and allowed to re-equilibrate for 10 min prior to the next injection. The eluate was sprayed directly into the ion source at the same flow rate. The capillary voltage was set at 3500 V and a cone gas flow of 15 L/h was used. Peptide sequencing was performed by data-directed analysis; the software was set to switch to MS/MS mode as soon as a doubly, triply or quadruply-charged ion was detected. For protein identification, the PKL files containing all the MS and MS/MS information obtained during each run were searched against the SWISS-PROT database (www.expasy.org) using Proteinlynx Global Server 1.1 (Waters). Protein identifications based on fragment ion data from one or two peptides only were manually verified before inclusion in our dataset of identified proteins.

2.6 Proteome informatics

We evaluated the potential biological significance of the differentially abundant proteins identified. To do this, we selected a group of proteins showing a volume ratio of $\geq$ 3 between the IDCA and normal tissue and that occured in at least 50% of the gels. Two types of researches were conducted. One of them was to study whether any of the selected proteins hasdbeen reported to be involved in signaling pathways, or in important protein-protein interactions. We employed the bioinformatics software, PathBlazer (InforMax, Frederick, MD, USA), loaded with Database of Interacting Proteins (DIP, http://dip.doe-mbi.ucla.edu), the Biomolecular Interaction Network Database (BIND, http://www.bind.ca) and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://fire2.scl.genome.ad.jp/kegg). We also searched the BioCarta website (www.biocarta.com) for signaling pathway information. Finally, we examined whether there are reported protein-protein interactions or codysregulations involving this group of differentially expressed proteins. For this purpose we developed a prototype bioinformatics tool using Access™ (Microsoft), to store and analyze the information retrieved from PubMed (http://www.ncbi.nlm.nih.gov/PubMed) using protein names as the query keywords, to search for co-occurrence of multiple proteins in the group.

3 Results

3.1 Cases

In this study we analyzed four breast tumors histologically classified by a licensed pathologist as IDCA, with the aim of determining if 2-D DIGE technologies will reveal proteins differentially expressed in IDCAs. The sizes of the tumors ranged from 0.5–7cm, tumor stage ranged from stage I to IIIA and age of the donors at diagnosis ranged from 30 y to 74 y (Table 1). All the tumors were poorly differentiated with a Scarf-Bloom classification index of 8 or 9. The ER and PR status of all the specimens were determined by immunohistochemistry (IHC) and the HER-2 status by fluorescence *in situ* hybridization (FISH). The protein sample used as reference for this comparative proteomic study, was an histologically normal breast tissue obtained from a 29 y old fully informed and consenting donor, with no personal or family history of breast cancer.

3.2 Protein abundance in infiltrating ductal carcinoma of the breast

All the samples studied were IDCAs with different tumor stage, grade and age of the donors. To maximize the total amount of protein loaded on each gel, we used whole sections from OCT embedded tissue rather than cells obtained by laser microdissection. A total of 175 μg of protein from each sample was used for 2-D DIGE as described in the Section 2.3. Four gels were run in parallel and each gel contained Cy dye-labeled proteins from one test sample (IDCA) and one reference sample (histologically normal). Gel images were generated with the Typhoon variable mode imager (Amersham Biosciences). Scanning at 10 μ resolution lasted for 10 min *per* channel and a minimum of 20 min was required to scan a two-color 2-D DIGE gel. The ImageQuant™ (Amersham Biosciences) file generated with the Typhoon imager was transferred to DeCyder for statistical analysis. Each pair of protein spots generated from the Cy3 (test) and Cy5 (reference) channels were converted to 3-D representations and displayed to show the relative peak volumes, height and area of each spot. The 3-D simulation of the protein spots allowed an objective inspection of spots corresponding to proteins from the Cy3 and Cy5 channels. By comparing each IDCA sample to a common normal (reference) sample, we were able to objectively estimate the abundance of similar proteins in IDCAs and generate quantitative data using the difference in-gel analysis (DIA) module in DeCyder.

Based on a threshold mode of $\geq$3.0, the DeCyder-DIA module detected protein spot differences that ranged from 14.8% in stage I IDCA to 30.6% in stage IIB IDCA (Table 2). As would be expected, 69–85% of the protein spots detected were not different in abundance between normal and cancer tissue, based on DIA. Our study detected differences in the overall abundance of proteins between the four IDCA specimens and normal tissue (Table 2). The mean of the population of increased proteins ranged from 6.47 in stage IIB IDCA, to 20.9 in stage IIIA IDCA, whereas the mean of the population of decreased proteins ranged from −5.2 in stage IIA IDCA to −6.0 in stage IIB IDCA (Table 2). We overlaid the images from the Cy5 and Cy3 channels and pseudo-colored each channel for better visual inspection. As shown in Fig. 1, the 2-D DIGE images are comparable but there are per-

Table 1. Histopathological properties of breast IDCAs analyzed by 2-D DIGE

	WRI[a)]-369	WRI[a)]-329	WRI[a)]-457	WRI[a)]-393
Histology	IDCA	IDCA	IDCA	IDCA
Tumor grade/Scarf-Bloom classification	Poorly Differentiated/8	Poorly Differentiated/9	Poorly Differentiated/9	Poorly Differentiated/8
Tumor stage (AJCC[b)] classification)	I	II A	II B	III A
Tumor size (cm)	0.5	1.2	4.2	7.0
Lymph node status (No. positive/total)	0/2	2/6	1/4	6/7
ER/PR status	−/−	−/−	−/+	+/+
HER-2 status/ratio (FISH[c)])	+/3.2	−/0.8	−/1.3	−/1.1
Age of donor (y)	75	52	47	30
Age at diagnosis (y)	74	52	46	30
Gender	F	F	F	F

a) WRI, Windber Research Institute
b) AJCC, American Joint Committee on Cancer
c) FISH, fluorescence *in situ* hybridization

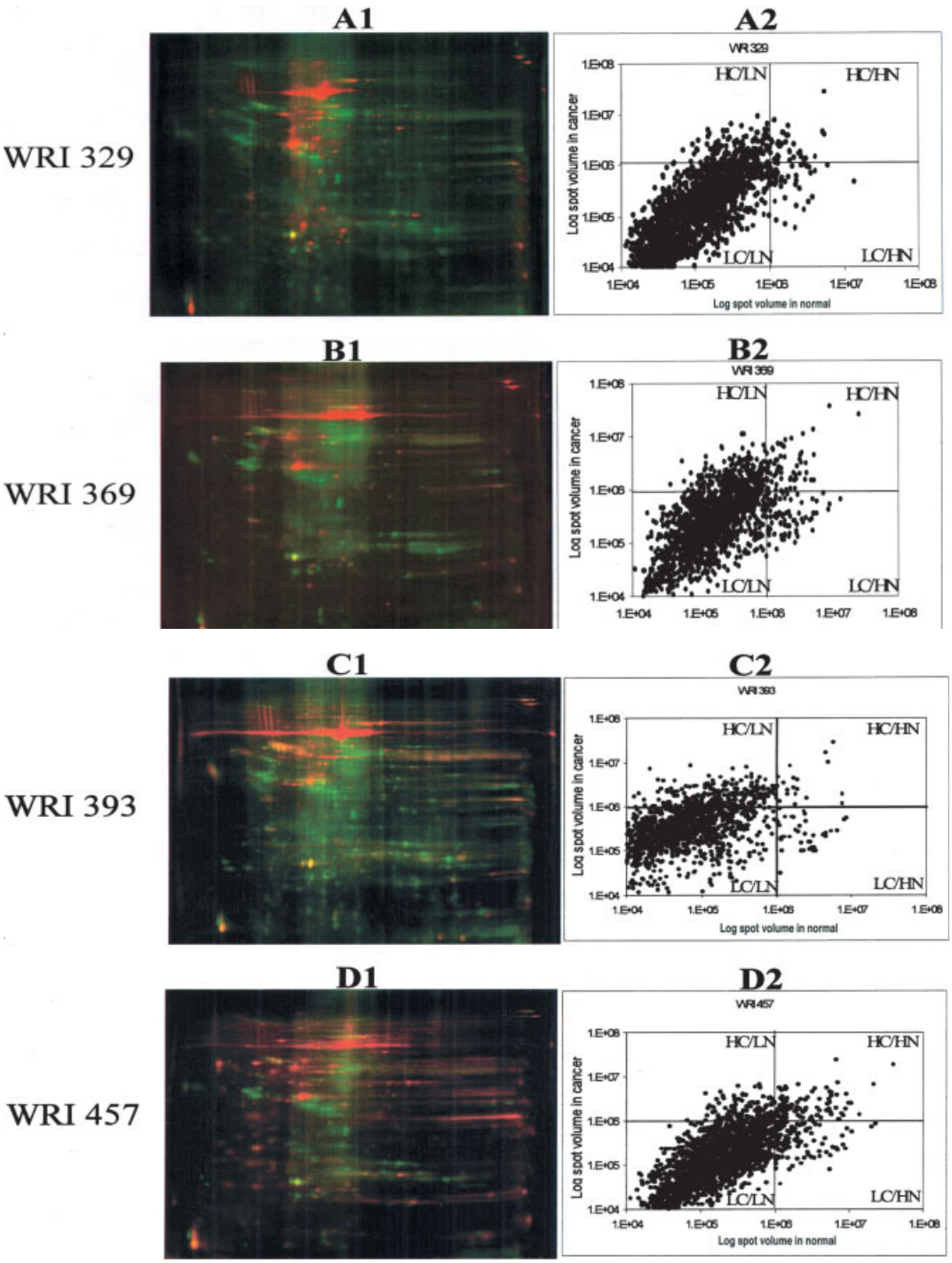

Figure 1. Two-color merged 2-D DIGE images in which the proteins expressed in infiltrating ductal carcinoma (WRI-329, A1; WRI-369, B1; WRI-393, C1; WRI-457, D1) were compared to proteins from normal (non-neoplastic) breast tisssue. Proteins from the carcinomas are colored red (Cy3) and proteins from the normal breast tissue are colored green (Cy5). Proteins that are similar in levels appear yellow. The scattergrams (A2, B2, C2, D2) to the right of the gels are plots of log spot volume in normal *vs.* log spot volume in IDCA. Each data point represents normalized spot volumes of the Cy3 channel (cancer, y-axis) and Cy5 channel (normal, x-axis) of all differentially abundant signals. The four quadrants represent, low in cancer and low-in-normal (LC/LN), low-in-cancer and high-in-normal (LC/HN), high-in-cancer and low–in-normal (HC/LN) and high-in-cancer and high-in normal (HC/HN). Most of the proteins detected were in the LC/LN quadrant.

Table 2. Global protein abundance in human breast IDCAs based on 2-D DIGE and DIA

Difference in-gel analysis statistics	Breast infiltrating ductal carcinomas			
	WRI[a)]-369	WRI[a)]-329	WRI[a)]-457	WRI[a)]-393
No. of proteins up-regulated in IDCA	25	84	139	74
No. of proteins unchanged in IDCA and normal	630	423	761	628
No. of proteins down-regulated in IDCA	84	60	196	75
Differentially expressed proteins (%)	109 (14.8)	144 (25.4)	335 (30.6)	149 (19.2)
Mean of increase population (± SD)	8.45 (8.0)	6.61 (6.2)	6.47 (6.2)	20.9 (36.4)
p-value – increase(*t*-test, 2-tailed, paired)	4.6×10^{-20}	5.1×10^{-9}	7.9×10^{-15}	6.0×10^{-5}
Mean of decrease population (± SD)	−5.5 (3.4)	−5.2 (2.7)	−6.0 (4.5)	−5.9 (5.3)
p-value – decrease (*t*-test, 2-tailed, paired)	1.0×10^{-29}	6.0×10^{-18}	6.5×10^{-22}	3.9×10^{-21}

a) Windber Research Institute

ceptible protein spot abundance and distribution differences as revealed by the log plot of the spot volumes of the IDCA *versus* the normal sample (Fig. 1). The majority of the protein spots were in the low-in-cancer (LC) and low-in-normal (LN) quadrant of all the four samples (Fig. 1). Summing up of all the normalized spot volumes and using the *t*-test (two tailed, paired) to compare the overall protein expression of all selected spots on the four gels demonstrated statistically significant differences between the IDCAs with *p*-values ranging from p = 1.03×10^{-29} to p = 6.0×10^{-5} (Table 2).

3.3 Protein identification by MS

A total of 737 protein spots showed $\geq$ three-fold difference (between test and reference) in at least one of the IDCAs. We selected 524 of such spots from the four gels (~130 spots/gel) for identification by MS. Out of the 524 protein spots picked and processed, ~80% (420/524) were unambiguously assigned identities by MALDI-MS or LC-MS/MS analysis. MALDI was employed as the first stage of mass spectrometric identification due to its high-throughput capability (288 samples in $<$ 24 h). Approximately 85% (357/420) of the proteins identified were assigned on the basis of MALDI peptide mass fingerprinting alone, and at least four peptides matched their *in silico* counterparts with an accuracy of $<$ = 0.1 Da. The remaining (~15%) were identified by LC-MS/MS with the benefit of the added peptide sequence information afforded by the data-directed analysis. The majority of the proteins that were picked but could not be identified by MALDI and LC-MS/MS analysis were found to be in the low-in-cancer (LC) and low-in-normal (LN) quadrants (Fig. 1) and so probably correspond to low abundant proteins occurring at levels below the minimum protein concentration required for in-gel digestion and detection with the mass spectrometers we used. Figure 2 shows an annotated representative gel (WRI-329); the location of some of the identified proteins and the corresponding list of proteins are shown in Table 3. The full list of proteins that displayed difference in abundance between normal breast tissue and IDCA can be found at our website (www.wriwindber.org/dige-idca/sir/03PE03/).

Three hundred and four of the proteins picked occurred on at least 50% of the gels and 41 proteins which displayed volume ratios of $\geq$ three-fold in at least one sample occurred on all four gels. Heat shock protein 70 (HSP70), proapolipoprotein, α-1-antitrypsin and peroxiredoxin-2 are the only proteins with $\geq$ three-fold difference between IDCA and normal that were found on the four gels, but the trend of abundance was not the same in all the samples. HSP70 and peroxiredoxin-2 were the only proteins that showed consistently higher and lower abundance, respectively, in all the IDCAs. Several of the proteins identified, *e.g.* serum albumin, alpha-1-antitrypsin, β-actin and PRO2619 were found at more than one location on the gels. Also, many proteins clustered in the mass and p*I* range of 38–72 kDa and 5.1–6.0, respectively (Fig. 2). We used *in silico* biology tools such as PathBlazer, a protein-protein interaction and pathway-building tool (Invitrogen-InforMax, Fredrick, MD, USA), to build pathways and identify proteins interacting with the proteins we selected (Table 3).

4 Discussion

This study was aimed at determining the feasibility of using the 2-D DIGE technology and a fully automated spot handling workstation for high-throughput proteomic analysis of human breast IDCA samples available at the CBCP-WRI tissue depository. 2-D DIGE has been

Table 3. Abundance (volume ratio) of selected proteins dysregulated in IDCA[a)]

Spot No.	Volume ratio[b)]	Protein ID	gi	Mass (Da)	p*I*
1	−4.75	Gelsolin	gi|121116	85679	6.2
2	4.93	Lumican	gi|20141464	38411	6.5
3	6.15	Serum albumin	gi|23307793	71334	6.1
4	16.74	Serum albumin	gi|23307793	71334	6.1
5	−3.27	Alpha-1-antitrypsin	gi|177831	46848	5.4
6	−3.13	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase)	gi|20070125	57480	4.76
7	−3.40	Alpha-1-antitrypsin	gi|177831	46718	5.5
8	−3.11	Alpha-1-antitrypsin	gi|1942629	44322	5.4
9	−6.02	Alpha-1-antitrypsin	gi|177831	46718	5.5
10	−9.98	PRO2619	gi|11493459	58513	5.96
11	5.95	Protein disulfide isomerase A3 (EC 5.3.4.1)	gi|2507461	57146	6
12	3.25	Fibrinogen beta chain	gi|399492	56577	8.54
13	−4.39	PRO2619	gi|11493459	58513	6
14	−8.19	PRO2619	gi|11493459	58513	6
15	3.68	Pyruvate kinase, M1 isozyme (EC 2.7.1.40)	gi|20178296	57787	8.2
16	−4.32	Alpha-1-antitrypsin	gi|177831	46718	5.5
17	−4.38	Alpha-1-antitrypsin	gi|177831	46718	5.5
18	−3.69	PRO2619	gi|11493459	58513	6
19	4.12	Fibrinogen gamma	gi|223170	46823	5.5
20	5.94	PRO2619	gi|11493461	58513	6
21	3.96	Actin, beta	gi|14250401	41341	5.6
22	4.61	Cytokeratin 18(424 AA)	gi|30311	47305	5.3
23	5.84	Actin, beta	gi|14250401	41321	5.56
24	−3.35	Similar to tubulin, beta 5	gi|12804891	38690	4.8
25	−4.97	Glycerol-3-phosphate dehydrogenase 1 (soluble)	gi|21594877	38171	5.8
26	3.16	Proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	gi|5453990	28876	5.8
27	−4.21	Carbonic anhydrase I: carbonic dehydratase	gi|4502517	28909	6.6

a) These proteins are shown in Fig. 2.
b) Volume ratio shows the fold difference between the test (IDCA, WRI329) and reference (normal) tissue. The negative sign (−) indicates that the protein occurred at a lower level in the IDCA.

used to study esophageal cancer and DeCyder™-DIA detected 58 proteins as up-regulated by > three-fold, 107 as down-regulated by > three-fold and 916 spots similar [11], but there were no reports on the use of this technology for proteomic analysis of breast IDCA. In-house evaluation studies of the 2-D DIGE technologies with breast cancer cell lines showed that ~98% reproducibility could be achieved if the recommended procedures are carefully followed (data not shown). As described in the Section 3.1, 2-D DIGE enabled the detection of qualitative and quantitative differences in protein abundance between pathologically different breast IDCAs. We detected differential protein abundance ranging from 14.8% (stage I IDCA) to 30.6% (stage IIB IDCA). Statistically four significant differences in the overall protein spot volumes were observed between the four IDCAs. This was not surprising because although all the samples were classified as IDCA, the ER, PR and HER-2

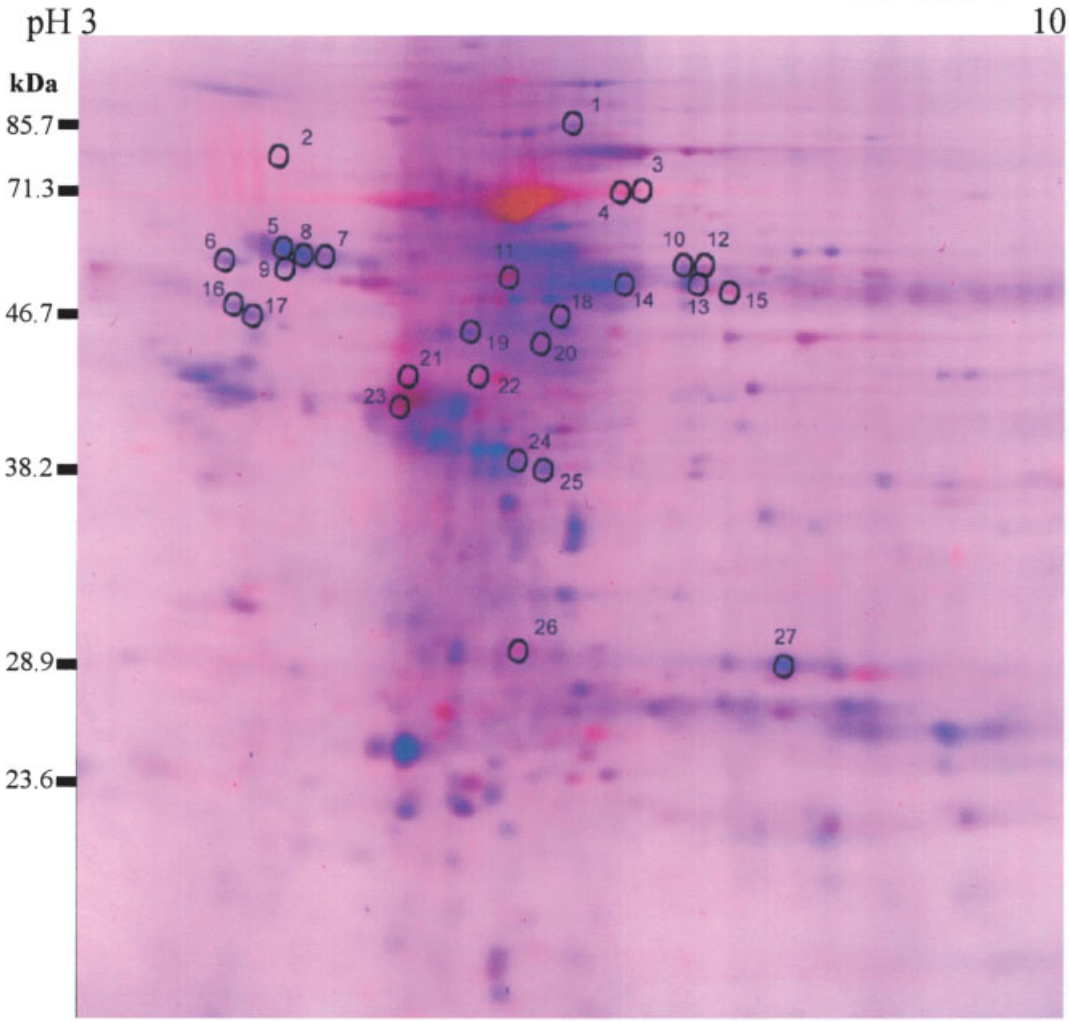

Figure 2. A representative 2-D DIGE image (WRI 329) showing 27 proteins found to be dysregulated (≥ three-fold) in breast IDCA. Proteins from the cancer tissue were labeled with Cy3 and those from the normal tissue were labeled with Cy5. Proteins that are less abundant in IDCA are colored blue (*e.g.* spot nos. 1, 5 and 9) and proteins that are more abundant in IDCA are colored red (*e.g.* spot nos. 2, 4 and 26). Proteins that are expressed at similar levels appear purple. The identities of these spots are presented in Table 3 (see Section 3.3).

status, histological stage, age at diagnosis, lymph node involvement and size were very different. Thus, the differences detected may be a true reflection of the global transcriptional/translational dysfunctions characteristic of different histological and pathological stages of IDCA. We used a relatively high cut-off (> three-fold difference) to select proteins for *in silico* biology analysis and expect that many of the selected proteins will also show the observed trend if the proteins from the same tissue sections are interrogated by Western analysis. This assumption is strengthened by experience in our laboratory with breast cancer cell lines and the report of Zhou and coworkers [11], which showed that proteins such as gp96, which showed more than three-fold difference between, test and reference samples after 2-D DIGE and DIA also showed differential expression by Western analysis.

The finding that the majority of the protein spots occurred in the low-in-cancer (LC) and low-in-normal (LN) quadrants was not surprising. This is because we utilized tissue sections from OCT specimens rather than whole tissue homogenates. This phenomenon, which is attributable to reduced tissue heterogeneity and/or low total protein loading, has been previously observed when proteomic analysis of breast tissue was performed with protein from tissue sections or cells captured by laser micro-

dissection were used [40]. A number of the differentially abundant proteins are known to be involved in breast cancer or other human diseases. For example, gelsolin, a protein that interacts with actin and regulates actin polymerization occurred at significantly lower levels in two of the IDCAs. Gelsolin and vinculin, also less abundant (~2.4-fold lower in IDCA) in 50% of the breast IDCAs, are involved in the Rho cell motility-signaling pathway and both are regulated by phosphatidyinositol-4-phosphate 5-kinase, type 1-alpha (see www.biocarta.com for the Rho cell motility pathway). Notably, β-actin is more abundant in 75% of the IDCA samples studied. This finding is of interest because firstly, β-actin is considered a house-keeping gene and its expression would be expected to be comparable in the test and reference samples. Secondly, gelsolin, which binds to β-actin, occurs at a lower level in stage II IDCAs studied. Gelsolin has been reported to be down-regulated in breast carcinoma and its down-regulation correlates with the progression to breast carcinoma [29, 30]. SWISS-PROT annotation also shows that dysfunction in gelsolin expression is responsible for "Finnish type" familial amyloidosis [29, 30]. Another protein involved in cell migration and proliferation that was identified as differentially abundant in IDCA is lumican (Fig. 2, Table 3). Lumican is a member of a small leucine-rich proteoglycan family (SLRP), which plays an important role during embryonic development, tissue repair and tumor growth [31]. Lumican genes and proteins have been found to be overexpressed in pancreatic cancer tissues, breast carcinoma, uterine cervical cancer cells, and in benign prostatic hyperplasia [31–34]. Furthermore, higher lumican expression was associated with higher tumor grade, lower estrogen receptor levels in the tumor, and younger age of the patients [35]. In this study, we noted higher levels of lumican (> four-fold) in stage IIA and IIB breast IDCA. In addition to the proteins described above, we also identified other differentially abundant proteins including pyruvate kinase, γ- and β-fibrinogen, cytokeratin 18, proline-4-hydroxylase and carbonic anhydrase (Table 3). Our use of a prototype literature-mining tool developed in-house, to mine reports in the public domain, also revealed literature that mentions some of the proteins we identified. For example, cDNA for cytokeratin-18, β-actin, and pyruvate kinase have been found to be expressed at high levels in pancreatic cancer cells but not in normal tissue [36]. Fibrinogen γ-chain and fibrinogen β-chain fragments were identified in various solid tumor types at the protein level, and fibrinogen γ-chain dimmer crosslinked by transglutaminase were detected in plasma from tumor patients but not in plasma from controls [37]. It is suggested that the elevation of β-fibrinogen correlates with tumor-associated fibrin deposition [37].

About 15% of the selected proteins could not be identified by peptide mass fingerprinting because (i) their concentration was too low, or (ii) there were no database hits with sufficient confidence. Based on our experience, we estimate that we can only assign identities to protein spots on 2-D DIGE gels equivalent to 50–80 fmol of protein. It will therefore be necessary to utilize more proteins (*i.e.* > 175 μg) if the visualization, automated spot handing and identification of such proteins by MS is desired. Our study revealed the presence of many proteins in the mass and p*I* cluster range of 38–72 kDa and 5.1–6.0, respectively. Performing IEF using narrow range [4–7] IPG gradients may allow these proteins to be more fully resolved and their extent of differential expression more accurately measured. In most cases, the calculated p*I* and molecular mass for the suggested protein identification reasonably matched the observed position of the protein feature; however in some cases the presence of proteins clearly distant from their apparent p*I* were observed. This may be partly due to protein comigration [41] because it has been reported that the p*I* of proteins are not expected to change since the dye molecule has a positive charge that compensates the charge on the lysine moiety [4]. The Cy dye adds ~ 0.5 kDa to the total M_r of the proteins [11], but this is generally not perceptible and did not cause any analytical problems.

There are concerns on the use of whole tissue for comparative genomics and proteomics studies because of tissue heterogeneity [39]. However, it is not always possible to get enough tissue to perform laser microdissection (LCM) and isolate enough cells for comparative proteomics research, without exhausting tissue that may be recalled for diagnostic purposes. To maximize the amount of protein loaded, we used whole sections, representing the tumor, rather than cells acquired by LCM. In spite of this, there were still protein spots detected by the DIA module of DeCyder that had concentrations that were apparently too low for automated in-gel digestion and mass spectrometric analysis. A number of the proteins picked and that showed differential abundance proved to be classical blood proteins. The prevalence of blood proteins like serum albumin and fibrinogen beta (Table 3) is due to the presence of blood capillaries that typically feed breast carcinomas. Although the use of tumor cells captured by LCM can reduce the abundance of blood proteins, these proteins were still present in the 2-D DIGE experiments we performed with proteins from cells procured by LCM [40].

In some cases, multiple spots of the same proteins were found at different locations on the gel. The proteins with multiple spots include PRO2619, β-actin and α-1-antitrypsin (Fig. 2, Table 3). PRO2619 is an albumin fragment, and more than five of the spots picked from different

regions of the same gel turned out to be PRO2619. Considering that PRO2619 is an albumin proteolysis fragment, it is likely that some protein breakdown may have occurred at the time of specimen collection or during the 2-DE process. On the other hand this phenomenon, which was also observed for vimentin, carbonic anhydrase, catalase and cytokeratin 8, may be partly due to protein comigration and/or the existence of multimolecular forms of the same protein. It is reported that since a maximum of ~3000 proteins can be observed on a typical 2-DE process, several proteins will inevitably comigrate to the same position [41] and isoforms of proteins or differentially modified forms (*e.g.* glycosylation) of the same protein may migrate to different locations on the gel. For this study, spots of the same protein at different locations of the gel were used for global spot volume analysis only if the protein spot occurred on all the four gels.

5 Concluding remarks

Although this is a preliminary study, it is apparent that 2-D DIGE has sufficient sensitivity as some of the proteins we identified have been reported in breast pathologies. This study demonstrates that 2-D DIGE is compatible with a fully automated spot handling workstation and by coupling *in silico* biology to wet laboratory experiments, it is possible to gain insights into the biochemical role played by dysregulated proteins prior to performing additional wet laboratory experiments. Further studies involving the use of more high quality IDCA samples are however required to verify the pattern of expression of the identified proteins.

We acknowledge the research funds and support received from the US Department of Defense, Henry Jackson Foundation for the Advancement of Military Medicine and Joyce Murtha Breast Care Center-Windber Medical Center for the Clinical Breast Care Project. The assistance received from Vimal Mittal, MD, Denise Papcunick and Brenda Deyarmin are also appreciated.

Received February 3, 2003

6 References

[1] Wulfkuhle, J. D., Mclean, K. C., Paweletz, C. P., Sgroi, D. C. *et al*., *Proteomics* 2001, *1*, 1205–1215.
[2] Steiner, S., Witsmann, F., *Electrophoresis* 2000, *21*, 2099–2104.
[3] Cellis, J., Kruhoffer, M., Gromova, I., Frederiksen, C. *et al*., *FEBS Lett.* 2000, *480*, 2–16.
[4] Westermeier, R, Naven, T., *Proteomics in Practice*, Weiley-VCH, Weinheim 2002, p. 316.
[5] Lewis, T. S., Hunt, J. B., Aveline, L. D., Jonscher, K. R. *et al*., *Mol. Cell* 2000, *6*, 1343–1354.
[6] Hanash, S.M, *Electrophoresis* 2000, *21*, 1202–1209.
[7] Bichsel, V. E., Liotta, L. A., Petricom, E. F. III., *Cancer J.* 2001, *7*, 69–78.
[8] Prasannan, L., Misek, D. E., Hinderer, R., Michon, J. *et al*., *Clin. Cancer Res.* 2000, *6*, 3949–3956.
[9] Liotta, L., Petricoin, E., *Nat. Rev. Genet.* 2000, *1*, 48–56.
[10] Emmert-Buck, M. R., Gillespie, J. W., Paweletz, C. P., Ornstein, D. K. *et al*., *Mol. Carcinog.* 2000, *27*, 158–165.
[11] Zhou, G., Li, H., DeCamp, D., Chen, S. *et al*., *Mol. Cell. Proteomics* 2002, *1*, 117–124.
[12] Wulfkuhle, J. D., Sgroi, D. C., Krutzsch, H., McLean, K. *et al*., *Cancer Res.* 2002, *62*, 6740–6749.
[13] Petricoin, E. F. III., Ardekani, M.A, Hitt, B. A., Levine, P. J. *et al*., *Lancet* 2002, *359*, 572–577.
[14] Sauter, E. R., Zhu, W., Fan, X-J, Wassell, R. P. *et al*., *Br. J. Cancer* 2002, *86*, 1440–1443.
[15] Hondermarck, H., Vercoutter-Edouart, A.-S., Revillion, F., Lemoine, J. *et al*., *Proteomics* 2001, *1*, 1216–1232.
[16] Kennedy, S., *Toxicol. Letts.* 2001, *120*, 379–384.
[17] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.
[18] Görg, A., Postel, W., Gunther, S., *Electrophoresis* 1988, *9*, 531–546.
[19] McDonough, J. L., Neverova, I., Van Eyk, J. E., *Proteomics* 2002, *2*, 978–987.
[20] Patton, W. F., *J. Chromatogr. B* 2002, *777*, 3–31.
[21] Stegemann, H., *Angew. Chem.* 1970, *82*, 640. (In German).
[22] Klose, J., *Humangenetik* 1975, *26*, 231–243.
[23] Bjellqvist, B., Sanchez, J. C., Pasquili, C, Ravier, F. *et al*., *Electrophoresis* 1993, *14*, 1375–1378.
[24] Rabilloud, T., *Electrophoresis* 1994, *15*, 278–282.
[25] von Eggeling, F., Gawriljuk, A., Fiedler, W., Ernst, G. *et al*., *Int. J. Mol. Med.* 2001, *8*, 373–377.
[26] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071–2077.
[27] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. *et al*., *Proteomics* 2001, *1*, 377–396.
[28] Gharbi, S., Gaffhey, P., Yang, A., Zvelebil, M. J. *et al*., *Mol. Cell. Proteomics* 2002, *7*, 91–98.
[29] Winston, J. S., Asch, H. L., Zhang, P. J., Edge, S. B. *et al*., *Breast Cancer Res. Treat.* 2001, *65*, 11–21.
[30] Asch, H. L, Winston, J. S., Edge, S. B., Stomper, P. C., Asch, B. B., *Breast Cancer Res. Treat.* 1999, *55*, 179–188.
[31] Ping Lu, Y., Ishiwata, T., Asano, G., *J. Pathol.* 2002, *196*, 324–330.
[32] Leygue, E., Snell, L., Dotzlaw, H., Troup, S. *et al*., *J. Pathol.* 2000, *192*, 313–320.
[33] Naito, Z., Ishiwata, T., Kurban, G., Teduka, K. *et al*., *Int. J. Oncol.* 2002, *20*, 943–948.
[34] Luo, J., Dunn, T., Ewing, C., Sauvageot, J. *et al*., *Prostate* 2002, *51*, 189–200.
[35] Leygue, E., Snell, L., Dotzlaw, H., Hole, K. *et al*., *Cancer Res.* 1998, *58*, 1348–1352.
[36] Paciucci, R., Berrozpe, G., Tora, M., Navarro, E. *et al*., *FEBS Lett.* 1996, *385*, 72–76.
[37] Gerner, C., Steinkellner, W., Holzmann, K., Gsur, A. *et al*., *Thromb. Haemost.* 2001, *85*, 494–501.
[38] Van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H. *et al*., *N. Engl. J. Med.* 2002, *347*, 1999–2009.
[39] Ellsworth, D. L., Shriver, C. D., Ellsworth, R. E., Deyarmin, B. *et al*., *Bio Techniques* 2003, *34*, 42–46.
[40] Sullivan, A. G., Russel, S., Arciero, C., Deyarmin, B. *et al*., *Proc. 50th ASMS Meeting on Mass Spectrometry and Allied Topics.* American Society of Mass Spectrometry, Orlando, Fl., June 2–6 2002, A20380.
[41] Lilley, K. S., Razzaq, A., Dupree, P., *Curr. Opin. Chem. Biol.* 2001, *6*, 46–50.