

The PCA and LDA analysis on the differential expression of proteins in breast cancer

Seng Liang^a, Manjit Singh^b, Saravanan Dharmaraj^a and Lay-Harn Gam^{a,*}

^a*School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang, Malaysia*

^b*Department of Surgery, Penang General Hospital, Penang, Malaysia*

Abstract. Breast cancer is a leading cause of mortality in women. In Malaysia, it is the most common cancer to affect women. The most common form of breast cancer is infiltrating ductal carcinoma (IDC). A proteomic approach was undertaken to identify protein profile changes between cancerous and normal breast tissues from 18 patients. Two protein extracts; aqueous soluble and membrane associated protein extracts were studied. Thirty four differentially expressed proteins were identified. The intensities of the proteins were used as variables in PCA and reduced data of six principal components (PC) were subjected to LDA in order to evaluate the potential of these proteins as collective biomarkers for breast cancer. The protein intensities of SEC13-like 1 (isoform b) and calreticulin contributed the most to the first PC while the protein intensities of fibrinogen beta chain precursor and ATP synthase D chain contributed the most to the second PC. Transthyretin precursor and apolipoprotein A-1 precursor contributed the most to the third PC. The results of LDA indicated good classification of samples into normal and cancerous types when the first 6 PCs were used as the variables. The percentage of correct classification was 91.7% for the originally grouped tissue samples and 88.9% for cross-validated samples.

Keywords: Breast cancer, proteomics, biomarkers, PCA, LDA

Abbreviations

IDC	infiltrating ductal carcinoma
2D-PAGE	two dimensional polyacrylamide gel electrophoresis
PCA	principal component analysis
LDA	linear discriminant analysis
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride
TCA	trichloroacetic acid
DTT	dithiothreitol
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
RT	room temperature
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
CBR-250	Coomassie Brilliant Blue 250
LC/MS/MS	liquid chromatography tandem mass spectrometry
ACN	acetonitrile

NH ₄ HCO ₃	ammonium bicarbonate
BSA	bovine serum albumin
PBS	phosphate buffered saline
HRP	horseradish peroxidase
4CN	4-chloro naphthol.

1. Introduction

Breast cancer is one of the major causes of death worldwide. It is responsible for 519 000 deaths around the world each year [1]. Breast cancer is the most common cancer affecting women, comprising 16% of all female cancer cases [2]. In Malaysia, breast cancer is the leading form of cancer in all ethnic and age groups for women and 1 in 19 women will have the disease in their lifetime [3]. Invasive ductal carcinoma is the most common form of invasive breast cancer, accounting for 85% of all breast cancer cases [4].

Proteomics is the study of the proteome, which is the entire complement of proteins expressed by an organism [5]. Proteomics approach has been used to identify protein biomarkers in breast cancer [6–11] and other types of cancers [12–15]. The functional roles of these

*Corresponding author: Dr. Lay-Harn Gam, School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia. Tel.: +604 6533888 ext 2208; Fax: +604 6570017; E-mail: layharn@usm.my.

Table 1
Patients information; ethnicity and age of patients as well as stage, grade and receptors status of the tissues

Patient no.	Race	Age	Stage	Grade	Estrogen receptor	Progesterone receptor	C-ERB-B ₂ oncoprotein
1	Indian	54	3	1	Positive	Positive	Negative
2	Chinese	67	2	3	Negative	Negative	Negative
3	Chinese	60	3	2	Positive	Positive	Positive
4	Malay	74	3	2	Positive	Positive	Negative
5	Malay	67	3	2	Negative	Negative	Positive
6	Malay	78	3	3	Positive	Positive	Negative
7	Chinese	64	3	3	Negative	Negative	Positive
8	Malay	63	2	3	Positive	Negative	Positive
9	Chinese	65	2	3	Negative	Negative	Negative
10	Chinese	59	2	3	Negative	Negative	Negative
11	Malay	55	4	3	Positive	Negative	Negative
12	Chinese	72	2	2	Positive	Negative	Negative
13	Chinese	80	3	2	Positive	Positive	Negative
14	Chinese	60	2	3	Negative	Negative	Negative
15	Malay	62	2	3	Negative	Negative	Positive
16	Indian	–	3	1	Positive	Positive	Positive
17	Indian	54	3	2	Negative	Negative	Negative
18	Indian	64	3	2	Positive	Positive	Negative

biomarkers in cancers may have important prognostic values in the detection and treatment of cancers [16, 17].

Principal component analysis (PCA) is a statistical test that analyzes multivariate data by reducing the number of variables for data analysis and interpretation [18]. PCA reduces the dimensionality of a multivariate data while retaining the most of the variance in the data. Linear discriminant analysis (LDA) is a technique to separate observations into several groups or categories [19]. The aim of LDA is to classify samples and identify variables that give optimum discrimination between the groups of test subjects. PCA has been used to analyze and reduce the dimensionality of proteome data in various experimental models [20–23], and together with LDA, they have been used to generate tumor classification models based on protein profile patterns in cancer [21].

In this study, we applied PCA and LDA on the differentially expressed aqueous soluble and membrane associated proteins between normal and cancerous breast tissues from breast cancer patients in Malaysia. The significance of these differentially expressed proteins as collective biomarkers for infiltrating ductal carcinoma among Malaysian female breast cancer was evaluated.

2. Methods and materials

2.1. Breast cancer patients

Eighteen breast cancer patients were included in the study. The patients were diagnosed with the disease for

the first time and had not received breast cancer pre-treatment. All patients were diagnosed with infiltrating ductal carcinoma (IDC) and had undergone surgery at Penang Hospital, Malaysia. Human ethical clearance for conducting this study was approved by the human ethical committee of Ministry of Health Malaysia and human ethical committee of Universiti Sains Malaysia, Malaysia. Informed consents were obtained from the patients before the tissues were collected for analysis. The information regarding patients is shown in Table 1.

2.2. Tissue collection

Tissue specimens of both normal and cancerous breast were collected at the Penang Hospital, Malaysia. In all cases, the normal breast tissues taken were adjacent to the cancerous parts. The tissue pathology was confirmed by the hospital's pathologist. Frozen sections of cancer tissue were taken from tissues in the anterior and deep region. Only cancerous tissues that contained greater than 90% malignant cells were used in this study. The breast tissues were stored at -80°C prior to analysis.

2.3. Protein extraction

Frozen breast tissues were thawed, rinsed with distilled water and cut into small pieces. Sequential extraction of aqueous soluble proteins and membrane associated proteins was performed based on the method by Molloy et al. [24]. Two volumes of TRIS buffer

(40 mM Tris, 1 mM AEBSF) were added to the tissues, which were homogenized in order to extract aqueous soluble proteins. The supernatant was recovered after centrifugation (13 000 rpm, 15 min, 20°C). The pellet was washed 3 times with 1 volume of TRIS buffer each before 1 volume of thiourea lysis buffer (TLB) [8 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.4% (w/v) carrier ampholytes, 50 mM DTT] was added to extract membrane associated proteins. The mixture was vortexed and the supernatant was recovered after centrifugation (13 000 rpm, 15 min, 20°C). The protein concentrations for both TRIS and TLB extracts were determined using RC-DC protein assay (Bio-Rad, USA). The aqueous soluble protein in TRIS extract was precipitated and concentrated using trichloroacetic acid (TCA) / acetone precipitation. The TRIS extract was incubated in 10% (w/v) TCA in ice-cold acetone containing 20 mM DTT at -20°C for 1.5 hours. After centrifugation (13 000 rpm, 15 min, 4°C) the supernatant was discarded. The pellet was washed in ice-cold acetone containing 20 mM DTT and centrifuged (13 000 rpm, 15 min, 4°C). After discarding the supernatant, the pellet was incubated in TLB for 1 hour to resolubilize the proteins. Each tissue samples was analyzed in duplicates.

2.4. 2D-PAGE

Protein extract containing 250 µg of protein was rehydrated into an 11 cm, pH 4–7 IPG strip (Bio-Rad, USA) for 15 hours at 20°C. Isoelectric focusing (IEF) was performed in a Protean IEF Cell (Bio-Rad, USA) at 20°C for 15 min at 250 V, 2.5 hours at 8000 V and sustained at 8000 V for 30000 V-hours. After IEF, the focused IPG strip was incubated in equilibration buffer I [6M Urea, 0.375M Tris-HCl, pH 8.8, 2% (w/v) SDS, 20% (v/v) glycerol, 1% (w/v) DTT] for 10 min at room temperature, followed by incubation in equilibration buffer II [6M Urea, 0.375M Tris-HCl, pH 8.8, 2% (w/v) SDS, 20% (v/v) glycerol, 2.5% (w/v) iodoacetamide] for 10 min at room temperature. The IPG strip was placed on top of a 10% sodium dodecyl sulfate polyacrylamide gel (16 × 13 × 0.1 cm) and electrophoresis was performed at 200 V for 3 hours according to the method of Laemmli using a Protean II xi Cell (Bio-Rad, USA). Proteins were visualized by staining the gel with Coomassie Brilliant Blue 250 (CBR-250) solution [0.1% (w/v) CBR-250, 40% (v/v) methanol, 10% (v/v) acetic acid] for 4 hours. The staining background was removed by washing the gel in a destaining solution [40% (v/v) methanol, 2% (v/v) acetic acid] twice for 2 hours each.

2.5. Gel imaging and analysis

Gel images were photographed using a Versadoc system (Bio-Rad, USA). The digitized gel images were processed using PDQuest software version 7.3 (Bio-Rad, USA). The software used the images to create a matchset to compare and match protein spots in every normal and cancerous breast tissues. Each protein spots was normalized as a percentage of the total density of each gel for quantitative purposes. A protein is up-regulated if its expression level in cancerous tissues is elevated by 1.5-fold or more as compared to normal breast tissue and was down-regulated when its expression level in the cancerous tissues was reduced by 1.5-fold or more as compared to the normal breast tissue. The statistical significance of the changes in protein expression was analyzed by the Wilcoxon-signed rank test in PDQuest.

2.6. In-gel digestion and LC-MS/MS analysis

In-gel digestion of protein spots was performed using the method by Othman et al. [25] Protein spots of interest were cut from the gel, washed with 100 mM ammonium bicarbonate (NH₄HCO₃) and dehydrated with acetonitrile (ACN). The protein was reduced and alkylated with incubating with DTT and iodoacetamide in 100 mM NH₄HCO₃ respectively. Trypsin was added to digest the proteins. The tryptic peptides were eluted from the gel pieces and dehydrated under constant flow of nitrogen gas. Thirty microlitres of 0.1% (v/v) formic acid in 85:15 solution of deionized water:acetonitrile were added to reconstitute the peptides and these were fractionated by RP-HPLC (C18, 150 X 0.3 mm, 5 µm 300Å) using an Agilent 1100 Series. The mobile phases A [0.1% (v/v) formic acid in deionized water] and B [0.1% (v/v) formic acid in ACN] were pumped at a constant flow rate of 4 µL/min. The peptides were eluted by linear gradient of 5% to 95% mobile phase B in 70 minutes and sustained at 95% mobile phase B for 5 min. The RP-HPLC was interfaced to an ESI-ion trap mass analyzer (Agilent). The two most intense ions in an MS scan that exceeded the set threshold (5000 counts) were isolated for MS/MS scans to produce series of product ions spectra for protein identification. The instrumental parameters used were: nebulizer pressure at 20 psi, auxiliary dry gas flow of 6 L/min, auxiliary dry gas temperature at 300°C, capillary voltage at 3.5 kV, exit capillary voltage 84.5 V, skimmer 1 voltage at 17.2 V, skimmer 2 voltage at 6.0 V. The MS scan region from 200–1800 m/z with

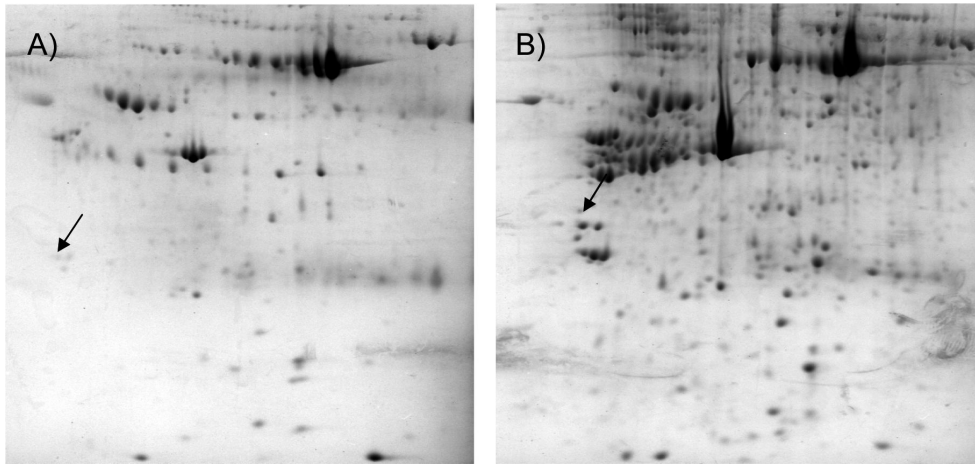


Fig. 1. 2D gel image of A) normal and B) cancerous breast tissue from Tris extract. The position of 14-3-3 zeta is marked by an arrow.

a scan time 1 s and an interscan time of 0.1 s were used. The parameters for MS/MS scan were: default collision energy (voltage) of 1.15V, charge state of 2, minimum threshold of 5000 counts, and isolation width of 2 m/z. Protein identification was done by subjecting the MS/MS data to a MASCOT (version 2.2) search using the MASCOT database search engine (MSDB), available at <http://www.matrixscience.com>. The search parameters used were: *Homo sapiens* for taxonomy, carboxymethyl for fixed modifications, peptide tolerance of ± 2 Da, MS/MS tolerance of ± 0.8 Da, average experimental mass value, one allowed miscleavage and unrestricted protein mass.

2.7. Western blot

Western blot was performed using the semi-dry method by Lauriere [26]. Proteins were separated by one-dimensional SDS-PAGE and incubated in cold transfer buffer (25 mM Tris, 192 mM glycine, 1.3 mM SDS, pH 8.3). Similar quantities of protein were transferred to a nitrocellulose membrane at 134 mA for 1.5 hours using a TE 70 Semiphor semi-dry transfer unit (Hoefer Scientific, Germany). The membrane was incubated with blocking buffer [3% (w/v) bovine serum albumin (BSA) in phosphate buffered saline (PBS), pH 7.2] for 2 hours at room temperature. The membrane was washed with washing buffer [0.1% (w/v) BSA, 0.1% (v/v) Tween 20 in PBS] and incubated in 20 mL of mouse anti-calreticulin monoclonal antibody (Abnova, Taiwan) with 1:1000 dilution in antibody diluent buffer [0.1% (w/v) BSA, 0.1% Tween 20, 0.02% sodium azide in PBS] overnight at room temperature. After washing, the membrane was incubated in 30 mL of

horseradish peroxidase (HRP) conjugated anti-mouse secondary antibody (Bio-Rad, USA) at 1:3000 dilution for 2 hours at room temperature. The membrane was then washed and after washing, the membrane was incubated in 20mL of 4-Chloro naphthol (4CN) substrate solution (Bio-Rad, USA) to visualize the protein band.

2.8. PCA and LDA

The protein spot intensities analyzed using PDQuest were imported to Statistical Package for Social Science (SPSS) version 15. SPSS was used to perform PCA based on protein spot intensities as variables. The PCs obtained from PCA were analyzed as variables in LDA using SPSS.

3. Results

Figure 1 shows the 2D gel images of Tris extracts of cancerous and normal tissues, while Fig. 2 shows the images of TLB extracts of the tissues. Figure 3a and b show the examples of spot intensities of 14-3-3 protein zeta and calreticulin extracted from 2D gel images in Tris and TLB buffer, respectively. The bar charts indicate the expression levels of the two proteins in normal and cancerous breast tissues. The differences in spots intensity of the same protein indicated that these proteins were differentially expressed between cancerous and normal tissues.

Table 2 lists the proteins identified in this study. The proteins' grand average of hydropathy (GRAVY) score indicates the hydrophilicity of the proteins, whereby the more negative the score, the higher is the hydrophilic

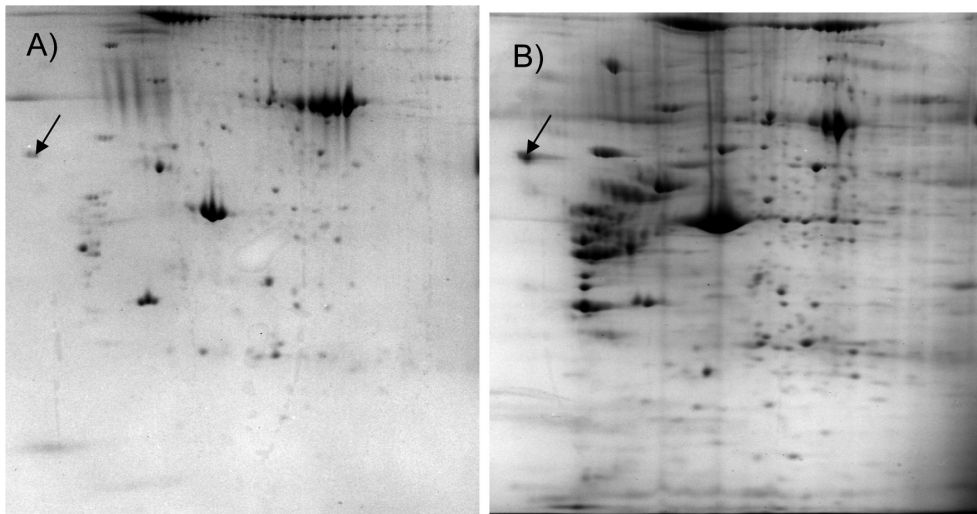


Fig. 2. 2D gel image of A) normal and B) cancerous breast tissue from TLB extract. The position of calreticulin is marked by an arrow.

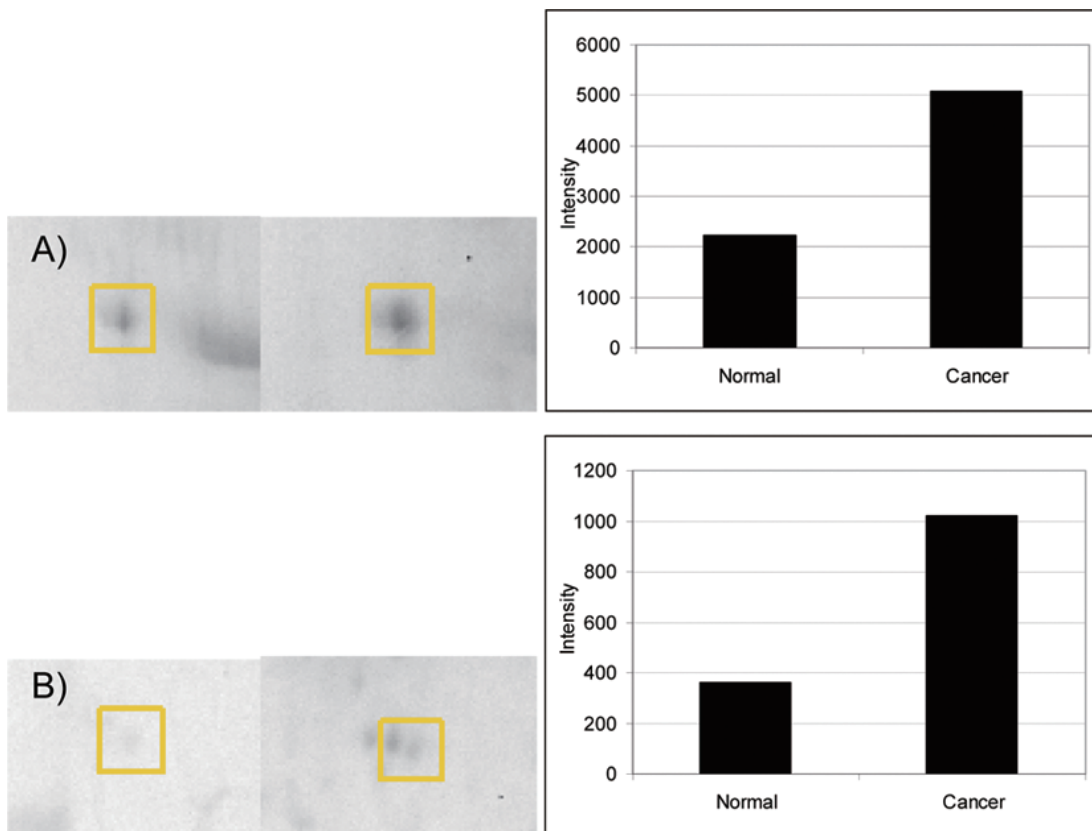


Fig. 3. Comparison of protein spot intensity between normal and cancerous breast tissue for A) Calreticulin and B) 14-3-3 protein zeta.

Table 2

Thirty-four differentially expressed proteins identified. Molecular information and MOWSE score as identified by MASCOT protein database by using the MS/MS data from mass spectrometry analysis. The total protein fold changed indicated levels of differential expression of proteins in cancerous and normal tissues

Protein spot no.	SwissProt accession number	Protein name	Molecular class	Molecular weight (Da)	pI	GRAVY score	MOWSE score (%)	Sequence coverage	Total fold change
1	P02768	Serum albumin precursor	Transport/cargo	71397	5.92	-0.354	96	14	+0.382
2	P00441	Superoxide dismutase	Oxidoreductase	16168	5.70	-0.344	52	9	+2.050
3	P32119	Peroxiredoxin-2	Oxidoreductase	21935	5.67	-0.199	260	26	+3.115
4	P00739	Haptoglobin-related precursor	Transport/cargo	39529	6.42	-0.308	43	3	-165.850
5	P02766	Transthyretin precursor	Transport/cargo	16003	5.52	-0.029	75	22	-37.355
6	P15090	Fatty acid binding protein	Carrier protein	14704	6.81	-0.249	223	24	-905.087
7	P00738	Haptoglobin precursor, allele 2 [validated]	Transport/cargo	45901	6.13	-0.421	56	6	-0.926
8	P02768	Serum albumin precursor [validated]	Transport/cargo	69366	5.92	-0.354	46	4	+69.451
9	P00739	Haptoglobin- related protein precursor	Transport/cargo	39529	6.42	-0.308	55	3	-53.381
10	P09211	Glutathione transferase	Transferase	23464	5.42	-0.121	277	53	-77.722
11	P68371	Class IV beta tubulin	Structural protein	50217	4.82	-0.362	52	13	341.172
12	P55735	SEC13-like 1, isoform b	Transport/cargo	36062	5.22	-0.372	115	37	511.402
13	P02675	Fibrinogen beta chain precursor	Coagulation factor	56624	8.54	-0.758	87	28	-673.998
14	P02675	Fibrinogen beta chain precursor	Coagulation factor	56624	8.54	-0.758	134	31	-544.111
15	P27797	Calreticulin	Calcium binding protein	47092	4.30	-1.104	107	29	4750.467
16	Not available	Unidentified protein	N/A	N/A	N/A	N/A	N/A	N/A	706.778
17	Q63610	Hypothetical protein	Hypothetical protein	27407	4.71	-0.992	59	27	1314.580
18	P63104	14-3-3 protein zeta (kinase regulator)	Adaptor molecule	27745	4.73	-0.621	193	47	1019.138
19	Q04917	14-3-3 protein eta	Adaptor molecule	28244	4.76	-0.618	71	23	2192.104
20	P52907	F-actin capping protein	Cytoskeletal protein	32965	5.45	-0.668	53	26	-46.854
21	P02766	Transthyretin precursor	Transport/cargo	16003	5.52	-0.029	85	22	709.938
22	P68133	Actin alpha	Cytoskeletal protein	38172	5.39	-0.161	64	17	-549.096
23	P07195	L-lactate dehydrogenase	Dehydrogenase	36928	5.71	0.056	222	34	+0.314
24	P21695	Glycerol-3- phosphate dehydrogenase	Dehydrogenase	38206	5.81	0.106	312	51	-3.292
25	P08758	Annexin V	Calcium binding protein	35994	4.94	-0.330	950	55	-155.125
26	P12111	Hypothetical protein DKFZp686K04147	Hypothetical protein	133297	5.42	0.029	57	3	-625.683
27	P02647	Apolipoprotein A-1 precursor	Transport/ cargo	30777	5.56	-0.717	368	47	-3326.52
28	O75947	ATP Synthase D chain	Hydrolase	18418	5.22	-0.554	80	60	-110.459
29	P32119	Peroxiredoxin-2	Oxidoreductase	21935	5.67	-0.199	376	42	+37.556
30	P10809	Heat shock protein 60	Heat shock protein	61212	5.70	-0.076	571	26	+374.259
31	P30101	Protein disulfide-isomerase	Isomerase	57202	5.98	-0.506	310	23	+292.646
32	P02675	Fibrinogen beta chain precursor	Coagulation factor	56624	8.54	-0.758	96	18	-173.312
33	P02675	Fibrinogen beta chain precursor	Coagulation factor	56624	8.54	-0.758	66	11	-249.408
34	P27797	Calreticulin	Calcium inding protein	47092	4.30	-1.104	207	14	+483.901

property. Although most of the proteins identified were up-regulated proteins in cancerous tissues, the expression levels of the same protein varied amongst patients in both cancerous and normal tissues. The total fold change referred to the relative accumulated intensity of the proteins between cancerous and normal tissues

from all the patients, where the positive sign indicated greater total intensity of the protein was identified in cancerous tissues while negative sign indicated greater total protein intensity was detected in the normal tissues.

Figure 4a shows the Western blot of calreticulin in

Table 3
Percentage of correct classification for LDA of normal and cancerous breast tissue samples

Samples	Number of correctly classified cases	Number of incorrectly classified cases	% Correct classification
Original grouped cases	33	3	91.7
Cross-validated cases	32	4	88.9

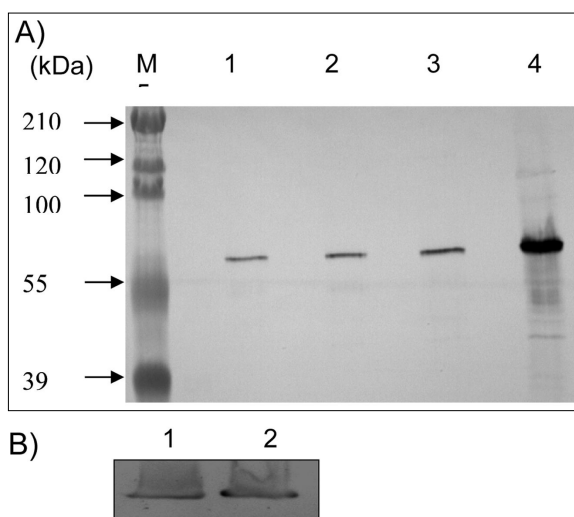


Fig. 4. Immunoblot of calreticulin. Lane M: Protein molecular weight markers (in kDa). Lane 1: Tris extract from normal tissue of the first patient. Lane 2: Tris extract from cancerous tissue of the first patient. Lane 3: Tris extract from normal tissue of the second patient. Lane 4: Tris extract from cancerous tissue of the second patient. B) The immunoblot of the positive control beta actin; Lane 1: Tris extract from normal tissue and Lane 2: Tris extract from cancerous tissue of the same patient.

both normal and cancerous protein extracts from two patients. The up-regulation of calreticulin was confirmed by the Western blot result. Figure 4b shows the Western blot result for beta-actin in normal and cancerous tissues, where the expression levels of beta-actin was found relatively similar between the two tissues. Beta-actin is used as a positive control in this study. The statistical significance of the changes in expression levels of these proteins was further tested using the Wilcoxon signed-rank test. The differential expressions of the proteins identified in this study were statistically significant at a 95% confidence level ($p < 0.05$). Figure 5 shows the percentage of up-regulation, equal expression, down-regulation and non-expression of the proteins in all patients.

Principal component analysis was performed on the normalized protein spot intensities for all the 34 protein spots identified. The number of variables was reduced to 10 PCs based on eigenvalues above one [27] and the first 6 PCs were used as variables in LDA. The first PC accounted for 28.12% of the total data variance, the

second PC accounted 12.45% and the third PC 8.32%. The fourth, fifth and sixth contributed respectively 7.76, 6.29 and 5.52% of the total variance. Overall, these 6 PC accounted for 68.46% of the total data variance.

The protein intensities of SEC13-like 1 (isoform b) and calreticulin contributed the most to the first PC while the protein intensities of fibrinogen beta chain precursor and ATP synthase D chain contributed the most to the second PC. Transthyretin precursor and apolipoprotein A-1 precursor contributed the most to the third PC. The protein intensities of SEC13-like 1 (isoform b) and calreticulin were the parameters that provide maximum discrimination between normal and cancerous breast specimens. Often, the first few PCs are associated with structure in the data or in this case differentiation between diseases states whereas the other PCs are associated with random variation [28]. The three-dimensional plot of the PCA is shown in Fig. 6, where the x-, y-, and z-axis were defined by the first, second and third PCs. The observations were separated into two regions.

The results of LDA using 6 PCs are shown in Table 3. The results indicated good classification of samples into normal and cancerous types when the first 6 PCs were used as the variables. Of all the original grouped samples, 91.7% were correctly classified while 88.9% of the cross-validated samples were correctly classified. In cross-validation, each sample is classified based on the characteristics of all the other remaining samples present.

4. Discussion

Normal and cancerous breast tissues obtained from human patients were used in the analyses, therefore, data obtained in this study represents the actual changes in protein expression between normal and cancerous tissues of infiltrating ductal carcinoma (IDC). Comparison was carried out in a pair-wise basis, where the pair of normal and cancerous tissues was obtained from the same patients. In the absence of laser capture microdissection [29], which is normally used for isolating specific cells of interest from microscopic regions of tissue by cutting the tissues directly with an ultraviolet

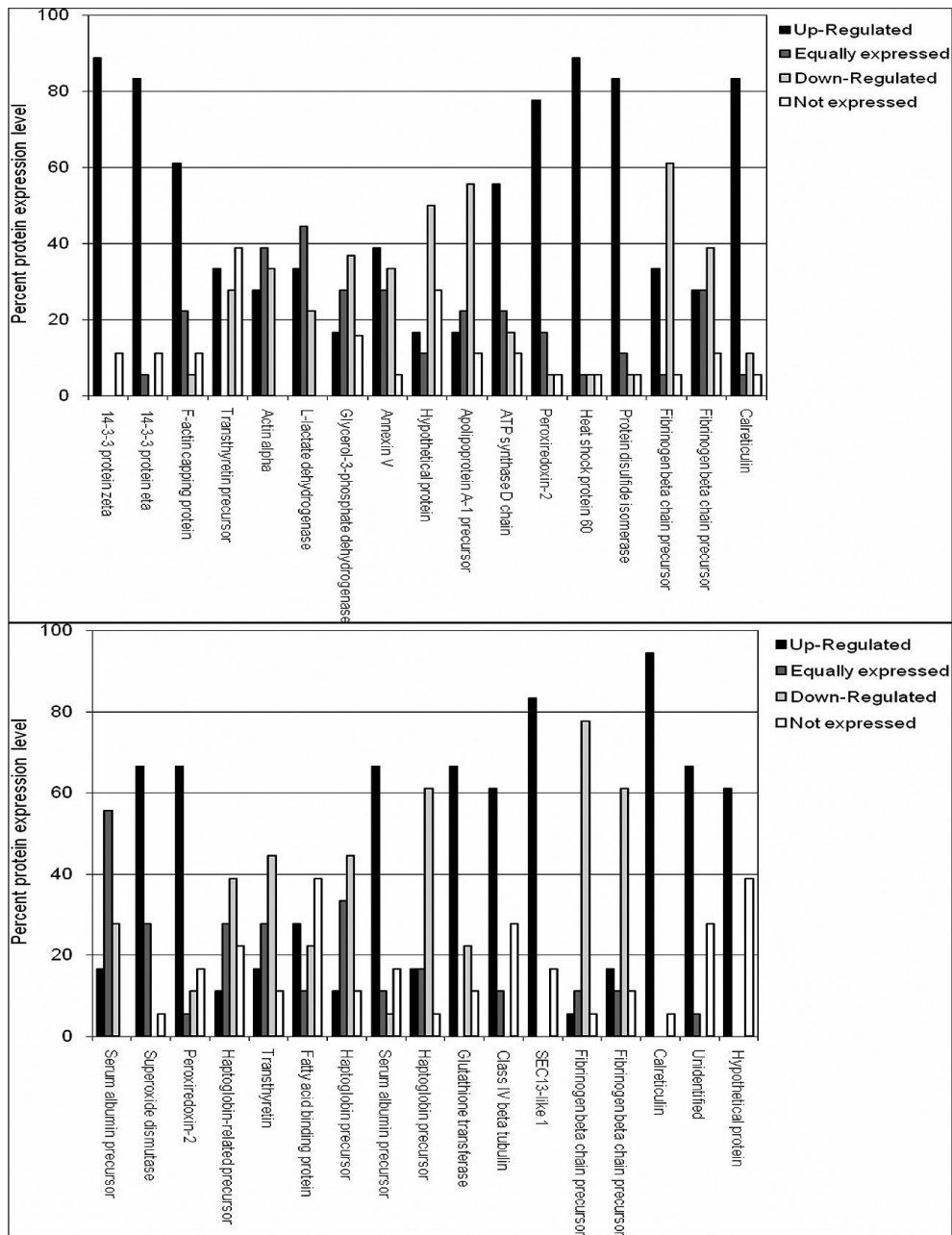


Fig. 5. Distribution of proteins in all patients.

pulsed laser, the potential biomarkers were identified through the determination of consistently differential expressed protein by cross-comparison between tissues from all the 18 patients tested. By using this approach, we minimized the possibility of identification of false proteins that were not related to IDC and at the same time, both epithelial-derived and stromal-derived proteins were extracted, where stromal-derived proteins

was recently suggested potentially serve as biomarkers of cancers [30].

The patients involved in this study were all afflicted with IDC and the selection was done without considering race, cancer stage, grade or the status of the receptors. There is no single protein detected as unique in either normal or cancerous breast tissue and there is also no single protein that was consistently (100%)

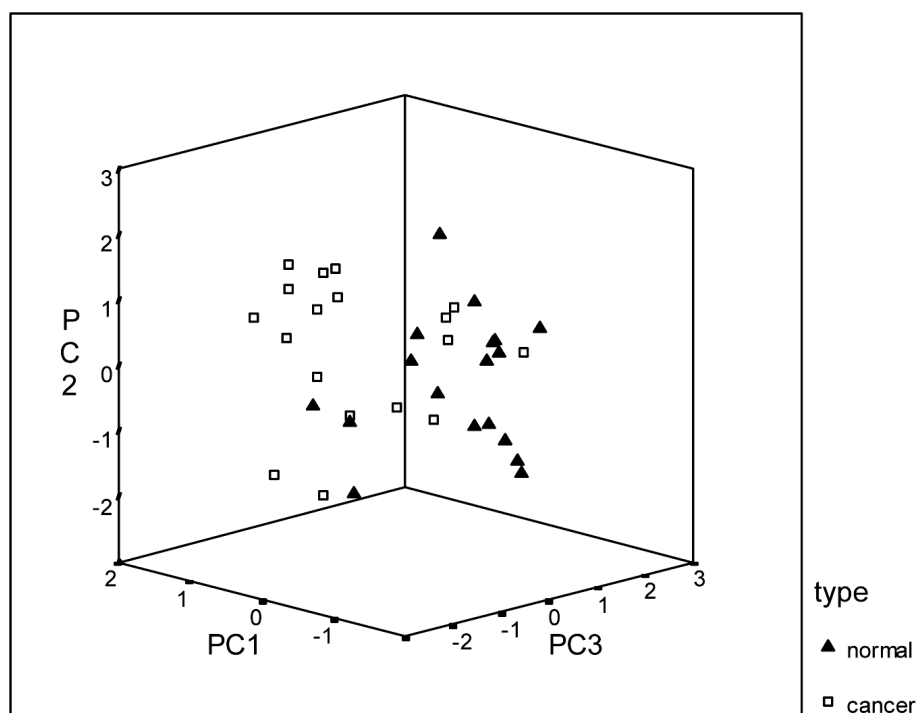


Fig. 6. Principal component score plot for normal and cancerous breast tissue samples.

up-regulated or down-regulated in cancerous or normal tissues, respectively. Due to tissues heterogeneity, the intensities of the proteins varied greatly between the patients although consistent expression levels (up-regulation and down-regulation) of the proteins can be detected in patients. Therefore, it is not possible to predict the state of IDC by referring only to one protein.

All the 34 differentially expressed proteins identified were subjected to PCA and the reduced data was used for LDA in order to determine the usability of these proteins for indicating the state of IDC. Our PCA and LDA had shown that the expression of six proteins out of the 34 proteins identified can be used reliably to predict the state of IDC, five of the proteins were up-regulated proteins; SEC13-like 1 (isoform b), calreticulin, ATP synthase D chain, Transthyretin precursor, apolipoprotein A-1 precursor. The only down regulated protein identified in this study capable of predicting IDC was fibrinogen beta chain precursor.

The original 34 variables (refer to the intensity of potential proteins identified) were reduced to 6 PCs that accounted for 68.46% of the total data variance, and these PCs were shown to be sufficient to indicate the state of cancer. This is shown by the subsequent LDA analysis which calculated that 91.7% of the originally grouped normal and cancerous tissues were correct-

ly classified while 88.9% of the cross-validated samples were correctly classified. This approach is used as LDA is sensitive to the ratio of sample size which is the number of observations to the number of variables. In general, five observations are recommended for each variable [31]. Moreover, using this approach, the parameters that are responsible for differentiation can be easily identified as they are the ones that contribute most to the reduced PCs. It is also noticed that 3 PCs are quite sufficient to show groupings as most of the breast normal or cancerous tissues are correctly grouped into defined regions in the PCA plot except for a few tissues that were not grouped correctly. This discrepancy could be caused by errors in sampling, variations in experimental conditions and individual patient variation.

Calreticulin is one of the two major proteins that contribute to the significance of PCA. There have been some previous reports on the up-regulation of this protein in cancers [32–35] and it had been marked as a marker for diagnosis of bladder cancer [36]. High expression of calreticulin in cancer has been suggested to prevent tumor cell lysis by stabilizing the cell membrane [37–39]. Calreticulin was detected in both Tris and TLB buffer extracts as up-regulated proteins in this study.

Protein SEC13 homolog (SEC13) is another major protein that contributes greatly to the PCA for discrimination of normal and cancerous breast tissues. SEC13 has not been reported previously for its involvement in cancer. It is a mammalian protein, which belongs to the SEC13 family of WD-repeat proteins. It is required for vesicle biogenesis from the endoplasmic reticulum during the transport of proteins [40] via budding of vesicles from the endoplasmic reticulum to form the Golgi apparatus [41].

The mitochondrial adenosine triphosphate synthase (ATPS), delta subunit is a subunit of the hydrolase enzyme ATPS that involve in synthesis of ATP for energy consumption of cells. ATPS is therefore crucial to the survival of cells. Down-regulation of ATP synthase beta subunit was reported in various types of cancer, including breast cancer [42,43], which was suggested to cause impairment of oxidative metabolism and therefore induce malignant cell growth [44,45]. In addition, inhibition of ATPS activity has been shown to inhibit the proliferation of several breast cancer cell lines by inducing apoptosis and arresting cell cycle [46–48]. In this study, ATPS was found up-regulated in cancerous tissues, and our finding is in agreement with Huang et al. [48].

Apolipoprotein A-1 (ApoA-1) is part of lipid-binding proteins family that constitutes the plasma lipoproteins. ApoA-1 precursor was found to be significantly over-expressed in IDC cells compared to normal breast cells [49]. The level of ApoA-1 in sera may hold a prognostic value in the individuals with high-risk breast cancer [50]. ApoA-1 is the major protein component of high density lipoprotein (HDL) that acts as a cofactor for the enzyme lecithin cholesterolacyltransferase (LCAT), which is responsible for the formation of plasma cholesteryl ester. In addition, it is involved in the transportation of cholesterol from tissues and arteries to the liver for excretion.

Transthyretin (TTR) or pre-albumin is carrier protein in serum that carries thyroid hormone thyroxine (T₄), retinol or Vitamin A. In addition, TTR was thought to sweep up toxic and foreign compounds in the blood stream. The increased levels of TTR are being used as a biomarker for the detection of colorectal carcinoma [51]. In contrary, TTR levels have been shown to decrease in patients with pancreatic cancer [52].

Fibrinogen beta chain (FBB) is the beta component of the fibrinogen glycoprotein, Following injury, fibrinogen is cleaved by the thrombin enzyme to form fibrin, which is the principal protein involved in the clotting of blood. FBB fragments were identified in

various solid types and may indicate cancer-related fibrin deposition [53,54]. The deposition of fibrin and fibrinogen into the extracellular matrix (ECM) of tumors serves to support the binding of growth factors and to promote the responses of cell proliferation, migration and adhesion during angiogenesis and tumor cell growth [55–57] and increase metastatic potential of circulating tumor cells [58]. In contrast, Chahed et al. [49] have reported that a particular type of fibrin and fibrinogen fragments that display anti-angiogenic effect was detected as down regulated protein in IDC. Fibrinogen was detected as a down-regulated protein with greater abundance of expression in the normal tissue as compared with the cancerous tissues, and its down-regulation features have shown to contribute to the identification of cancerous tissues.

5. Conclusion

The patients involved in this study comprised of combination of Chinese, Indian and Malay whom were diagnosed with different stages, grades of IDC breast cancer, and with the combination of different receptors, the LDA indicated that the classification of cancerous and normal tissues by collective used of the biomarkers can be performed at high accuracy levels regardless of grade, stage and receptors status of the tissues, and race of the patients. PCA of the whole data set reduced the data to ten PCs of which six PCs gave reliable discrimination of the cancerous and normal tissue. Six proteins that showed high standard deviations between patients had high contributions to the first three PCs and these proteins were associated with discrimination of both diseased and non-diseased samples. We see the possibility of using this similar approach in prediction the prognosis of the disease, which we will study in the near future.

Acknowledgments

We would like to thank the Ministry of Science, Technology and Innovation Malaysia for providing a grant (Project number: 02-01-05-SF0353) to fund this project, National Institute of Pharmaceutical and Nutraceutical, Ministry of Science, Technology and Innovation Malaysia for the use of the mass spectrometry for protein analysis, the Ministry of Health, Malaysia for providing the tissues and human ethical clearance for conducting this research.

References

- [1] World Health Organization, Cancer: Fact Sheet, [updated February 2009], 2009. Available from: <http://www.who.int/entity/mediacentre/factsheets/fs297/en/index.html>.
- [2] World Health Organization, Breast Cancer Burden, 2010. Available from: <http://www.who.int/cancer/detection/breastcancer/en/index1.html>.
- [3] Malaysian Oncological Society, Breast Cancer, 2010. Available from: <http://www.malaysiaoncology.org/article.php?aid=114>.
- [4] R.B. Dickson, R.G. Pestell and M.E. Lippman, Cancer of the breast, in: *Cancer – Principals and Practice of Oncology*, V.T. Devita and S.A. Rosenberg, eds, Lippincott Williams and Wilkins, Philadelphia, 2005, pp. 1399–1487.
- [5] M.R. Wilkins, C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.C. Sanchez, J.X. Yan, A.A. Gooley, G. Hughes, I. Humphrey-Smith, K.L. Williams and D.F. Hochstrasser, From protein to proteomics: large-scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Biotechnology* **14** (1996), 61–65.
- [6] C.S. Giometti, S.L. Tollaksen, C. Chubb, C. Williams and E. Huberman, Analysis of proteins from human breast epithelial cells using two-dimensional gel electrophoresis, *Electrophoresis* **16** (1995), 1215–1224.
- [7] S. Liang, M. Singh and L.H. Gam, The differential expression of aqueous soluble proteins in breast normal and cancerous tissues in relation to ethnicity of the patients; Chinese, Malay and Indian, *Disease Markers* **28** (2010), 149–165.
- [8] F. Le Naour, D.E. Misk, M.C. Krause, L. Deneux, T.J. Giordano, S. Scholl and S.M. Hamash, Proteomics-based identification of RS/DJ-1 as a novel circulating tumour antigen in breast cancer, *Clin Cancer Res* **7** (2001), 3325–3327.
- [9] D.Q. Li, L. Wang, F. Fei, Y.F. Hou, J.M. Luo, Wei Chen, R. Zeng, J. Wu, J.S. Lu, G.H. Di, Z.L. Ou, Q. Xia, Z.Z. Shen and Z.M. Shao, Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry, *Proteomics* **6** (2006), 3352–3368.
- [10] M.J. Page, B. Amess, R.R. Townsend, R. Parekh, A. Herath, L. Brusten, M.J. Zvebil, R.C. Stein, M.D. Waterfield, S.C. Davies and M.J. O'Hare, Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties, *Proc Natl Acad Sci USA* **96** (1999), 12589–12594.
- [11] K. Williams, C. Chubb, E. Hubermann and C.S. Giometti, Analysis of differential protein expression in normal and neoplastic human breast epithelial cell lines, *Electrophoresis* **19** (1998), 333–343.
- [12] P. Alfonso, A. Nuñez, J. Madoz-Gurpide, L. Lombardia, L. Sanchez and J.I. Casal, Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis, *Proteomics* **5** (2005), 2602–2611.
- [13] R. Chen, S. Pan, T.A. Bretnall and R. Aebersold, Proteomic profiling of pancreatic cancer for biomarker discovery, *Mol Cell Proteomics* **4** (2005), 523–533.
- [14] D.B. Friedman, S. Hill, J.W. Keller, N.B. Merchant, S.E. Levy, R.J. Coffey and R.M. Caprioli, Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry, *Proteomics* **4** (2004), 793–811.
- [15] P.R. Jungblut, U. Zimny-Arndt, E. Ziehl-Eberhart, J. Stulík, K. Koupilova, K.P. Pleissner, A. Otto, E.C. Müller, W. Sokolowska-Köhler, G. Grabher and G. Stöfler, Proteomics in human disease: cancer, heart and infectious diseases, *Electrophoresis* **20** (1999), 2100–2110.
- [16] M.R. Emmert-Buck, J.W. Gillespie, C.P. Pawaletz, D.K. Ornstein, V. Basrur, E. Appella, Q.H. Wang, J. Huang, N. Hu, P. Taylor and E.F. Petricoin, An approach to proteomic analysis of human tumors, *Mol Carcinog* **27** (2000), 158–165.
- [17] F.J. Esteva and G.N. Hortobagyi, Prognostic molecular markers in early breast cancer, *Breast Cancer Res* **6** (2004), 109–118.
- [18] M.J. Karson, *Multivariate statistical methods: An introduction*, Iowa State University Press, Iowa, 1982, p. 159, 191.
- [19] N.C. Giri, *Multivariate Statistical Analysis*, Marcel Dekker, New York, 1996, p. 293.
- [20] S.O. Deininger, M.P. Ebert, A. Fütterer, M. Gerhard and C. Röcken, MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers, *J Proteome Res* **7** (2008), 5230–5236.
- [21] M.C. Djidja, E. Claude, M. Snel, S. Francese, P. Scriven, V. Carolan and M. Clench, Novel molecular tumour classification using MALDI–mass spectrometry imaging of tissue microarray, *Anal Bioanal Chem* **397** (2010), 587–601.
- [22] J.M. Hillegass, A. Shukla, M.B. MacPherson, J.P. Bond, C. Steele and B.T. Mossman, Utilization of gene profiling and proteomics to determine mineral pathogenicity in a human mesothelial cell line (LP9/TERT-1), *J Toxicol Environ Health A* **73** (2010), 423–436.
- [23] E. Marengo, E. Robotti, M. Bobba and F. Gosetti, The principle of exhaustiveness versus the principle of parsimony: a new approach for the identification of biomarkers from proteomic spot volume datasets based on principal component analysis, *Anal Bioanal Chem* **397** (2010), 25–41.
- [24] M.P. Molloy, B.R. Herbert, B.J. Walsh, M.I. Tyler, M. Traini, J.C. Sanchez, D.F. Hochstrasser, K.L. Williams and A.A. Gooley, Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis, *Electrophoresis* **19** (1998), 837–844.
- [25] M.I. Othman, M.I. Majid, M. Singh, S. Subathra, L. Seng and L.H. Gam, Proteomics of grade 3 infiltrating ductal carcinoma in Malaysian Chinese breast cancer patients, *Biotechnol Appl Biochem* **52** (2009), 209–219.
- [26] M. Laurière, A semidry electroblotting system efficiently transfers both high- and low-molecular weight proteins separated by SDS-PAGE, *Anal Biochem* **212** (1993), 206–211.
- [27] S. Sharma, *Applied multivariate techniques*, John Wiley and Sons, New York, 1996, pp. 241–244.
- [28] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers-Verbeke, eds, *Handbook of Chemometrics and Qualimetrics: Part A, Volume 20, Part 1*, Elsevier Science, 1998.
- [29] M.R. Emmert-Buck, R.F. Bonner, P.D. Smith, R.F. Chuaqui, Z. Zhuang, S.R. Goldstein, R.A. Weiss and L.A. Liotta, Laser capture microdissection, *Science* **274** (1996), 998–1001.
- [30] A.S. Felix, J. Weissfeld, R. Edwards and F. Linkov, Future directions in the field of endometrial cancer research: the need to investigate the tumor microenvironment, *Eur J Gynaecol Oncol* **31** (2010), 139–144.
- [31] J.J.F. Hair, R.E. Anderson, R.L. Tatham and W.C. Black, *Multivariate data analysis*, Prentice Hall, New Jersey, 1998, 245, 258, 260–262.
- [32] L. Bini, B. Magi, B. Marzocchi, F. Arcuri, S. Tripodi, M. Cintonino, J.C. Sanchez, S. Frutiger, G. Hughes, V. Pallini, D.F. Hochstrasser and P. Tosi, Protein expression profiles in hu-

- man breast ductal carcinoma and histologically normal tissue, *Electrophoresis* **18** (1997), 2832–2841.
- [33] B. Franzén, S. Linder, A.A. Alaiya, E. Eriksson, K. Fujioka, A.C. Bergman, H. Jörnvall and G. Auer, Analysis of polypeptide expression in benign and malignant human breast lesions, *Electrophoresis* **18** (1997), 582–587.
- [34] K. Chahed, M. Kabbage, B. Hamrita, C.L. Guillier, M. Trimeche, S. Remadi, L. Ehret-Sabatier and L. Chouchane, Detection of protein alterations in male breast cancer using two dimensional gel electrophoresis and mass spectrometry: The involvement of several pathways in tumorigenesis, *Clin Chim Acta* **388** (2008), 106–114.
- [35] H.A. Sarvajia, J.H. Yoon and I.M. Lazar, Proteome profile of the MCF7 cancer cell line: A mass spectrometric evaluation, *Rapid Commun Mass Spectrom* **20** (2006), 3039–3055.
- [36] S. Kageyama, T. Isono, H. Iwaki, Y. Wakabayashi, Y. Okada, K. Kontani, K. Yoshimura, A. Terai, Y. Arai and T. Yoshiki, Identification by proteomic analysis of calreticulin as a marker for bladder cancer and evaluation of the diagnostic accuracy of its detection in urine, *Clin Chem* **50** (2004), 857–866.
- [37] S.A. Fraser, M. Michalak, W.H. Welch and D. Hudig, Calreticulin, a component of the endoplasmic reticulum and of cytotoxic lymphocyte granules, regulates perforin-mediated lysis in the hemolytic model system, *Biochem Cell Biol* **76** (1998), 881–887.
- [38] S.A. Fraser, R. Karimi, M. Michalak and D. Hudig, Perforin lytic activity is controlled by calreticulin, *J Immunol* **164** (2000), 4150–4155.
- [39] Z.D. Juranic, Z. Neskovic-Konstantinovic, T.P. Stanojkovic, Z. Zizak, T. Srdic, N. Stanojevic-Bakic, D. Milosevic and D. Jovanovic, The antitumor immune response in HER-2 positive, metastatic breast cancer patients, *J Transl Med* **3** (2005), 13.
- [40] D.A. Shaywitz, L. Orci, M. Ravazzola, A. Swaroop and C.A. Kaiser, Human SEC13Rp functions in yeast and is located on transport vesicles budding from the endoplasmic reticulum, *J Cell Biol* **128** (1995), 769–777.
- [41] B.L. Tang, F. Peter, J. Krijnse-Locker, G. Griffiths and W. Hong, The mammalian homology of yeast Sec13p is enriched in the intermediate compartment and is essential for protein transport from the endoplasmic reticulum to the Golgi apparatus, *Mol Cell Biol* **17** (1997), 256–266.
- [42] J.M. Cuezva, M. Krajewska, M.L. de Heredia, S. Krajewski, G. Santamaria, H. Kim, J.M. Zapata, H. Marusawa, M. Chamorro and J.C. Reed, The bioenergetic signature of cancer: A marker of tumor progression, *Cancer Res* **62** (2002), 6674–6681.
- [43] A. Isidoro, M. Martinez, P.L. Fernández, A.D. Ortega, G. Santamaria, M. Chamorro, J.C. Reed and J.M. Cuezva, Alteration of the bioenergetic phenotype of mitochondria is a hallmark of breast, gastric lung and oesophageal cancer, *Biochem J* **378** (2004), 17–20.
- [44] O. Warburg, On the origin of cancer cells, *Science* **123** (1956), 309–314.
- [45] T.J. Schulz, R. Thierbach, A. Voigt, G. Drewes, B. Mietzner, P. Steinberg, A.F.H. Pfeiffer and M. Ristow, Induction of oxidative metabolism by mitochondrial frataxin inhibits cancer growth: Otto Warburg revisited, *J Biol Chem* **281** (2006), 977–981.
- [46] D.R. Green and J.C. Reed, Mitochondria and apoptosis, *Science* **281** (1998), 1309–1312.
- [47] S. Matsuyama, Q. Xu, J. Velours and J.C. Reed, The mitochondrial F0F1-ATPase proton pump is required for function of the proapoptotic protein Bax in yeast and mammalian cells, *Mol Cell* **1** (1998), 327–336.
- [48] T.C. Huang, H.Y. Chang, C.H. Hsu, W.H. Kuo, K.J. Chang and H.F. Juan, Targeting therapy for breast carcinoma by ATP synthase inhibitor Aurovertin B, *J Proteome Res* **7** (2008), 1433–1444.
- [49] K. Chahed, M. Kabbage, L. Ehret-Sabatier, C. Lemaitre-Guillier, S. Remadi, J. Hoebeke and L. Chouchane, Expression of fibrinogen E-fragment and fibrin E-fragment is inhibited in the human infiltrating ductal carcinoma of the breast: The two-dimensional electrophoresis and MALDI-TOF mass spectrometry analyses, *Int J Oncol* **27** (2005), 1425–1431.
- [50] A. Gonçalves, B. Esterni, F. Bertucci, R. Sauvan, M. Cubizoles, S. Granjeaud, C. Chabannon, G. Houvenaegel, J. Jacquemier, X.Y. Meng, E.T. Fung, D. Dirnbaum, D. Maranchi, P. Vines and J.P. Borg, Postoperative serum proteomic profiles and identification of biomarkers with prognosis value in high-risk early breast cancer patients, *Breast Cancer Res* **7** (2005), P05.03.
- [51] A.K. Fentz, M. Spörl, J. Spangenberg, H.J. List, C. Zornig, A. Dörner, P. Layer, H. Juhl and K.A. David, Detection of colorectal adenoma and cancer based on transthyretin and C3adesArg serum levels, *Proteomics Clin Appl* **1** (2007), 536–544.
- [52] M. Ehmann, K. Felix, D. Hartmann, M. Schnölzer, M. Nees, S. Vorderwülbecke, R. Bogumil, M. Büchler and H. Friess, Identification of potential markers for the detection of pancreatic cancer through comparative serum protein expression profiling, *Pancreas* **34** (2007), 205–214.
- [53] C. Gerner, W. Steinkellner, K. Holzmann, A. Gsur, R. Grimm, C. Ensinger, P. Obrist and G. Sauermaier, Elevated plasma levels of crosslinked fibrinogen gamma-chain dimer indicate cancer-related fibrin deposition and fibrinolysis, *Thromb Haemost* **85** (2001), 494–501.
- [54] R.I. Somiari, S. Somiari, S. Russell and C.D. Shriver, Proteomics of breast carcinoma, *J Chromatogr B* **815** (2005), 215–225.
- [55] P.J. Simpson-Haidaris and B. Rybarczyk, The role of fibrinogen as an extracellular matrix protein, *Ann NY Acad Sci* **936** (2001), 406–425.
- [56] C.G. Colpaert, P.B. Vermeulen, I. Benoy, A. Soubry, F. van Roy, P. van Beest, G. Goovaerts, I.Y. Dirix, P. van Dam, S.B. Fox, A.L. Harris and E.A. Marck, Inflammatory breast cancer shows angiogenesis with high endothelial proliferation rate and strong E-cadherin expression, *Br J Cancer* **88** (2003), 718–725.
- [57] A. Sahni, P.J. Simpson-Haidaris, S.K. Sahni, G.G. Vaday and C.W. Francis, Fibrinogen synthesized by cancer cells augments the proliferative effect of fibroblast growth factor-2 (FGF-2), *J Thromb Haemost* **6** (2008), 176–183.
- [58] J.S. Palumbo, K.W. Kombrinck, A.F. Drew, T.S. Grimes, J.H. Kiser, J.L. Degen and T.H. Bugge, Fibrinogen is an important determinant of the metastatic potential of circulating tumor cells, *Blood* **96** (2000), 3302–3309.