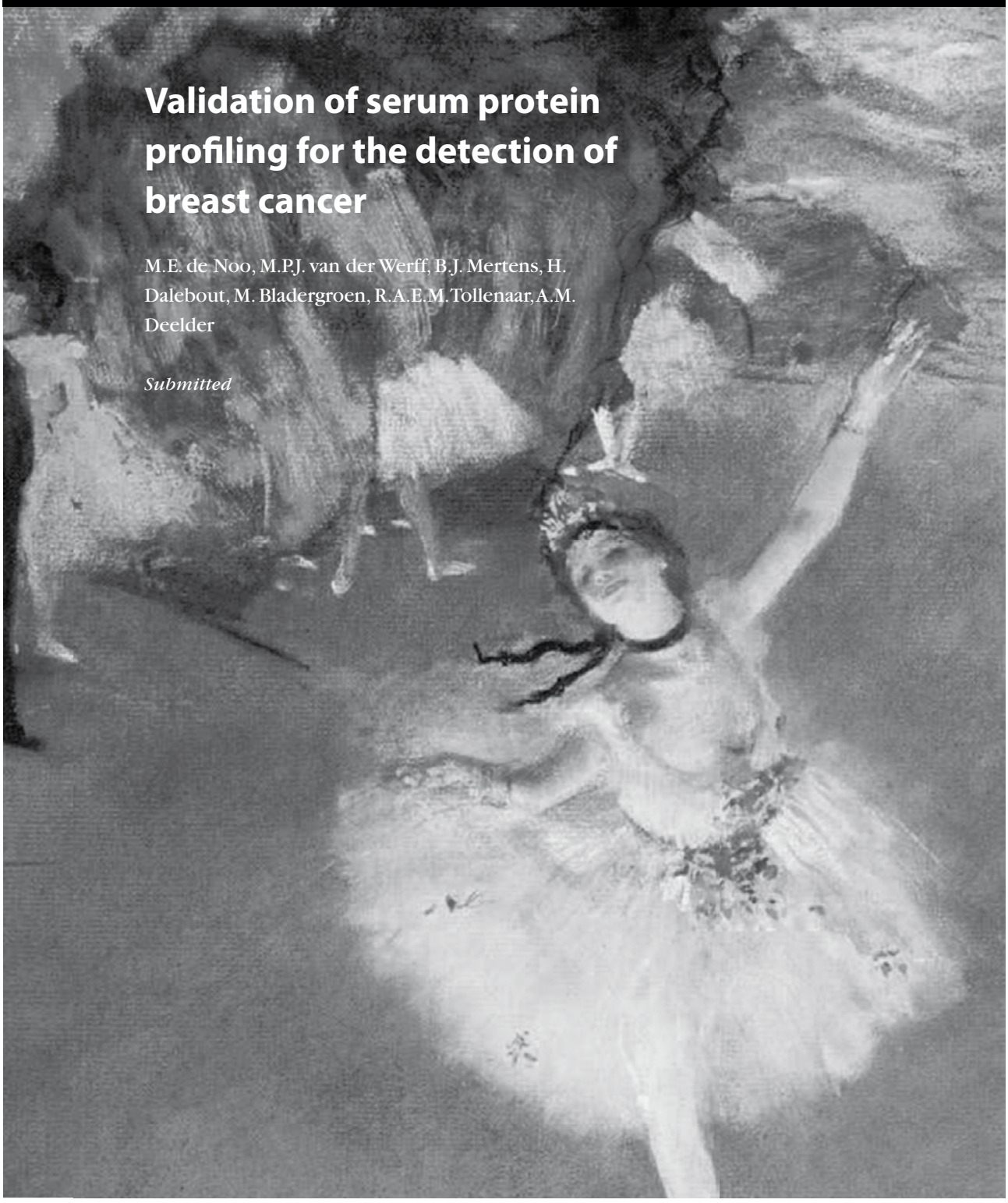
Clinical proteomics in oncology : a passionate dance between science and clinic

Noo, M.E. de

Chapter 7



Validation of serum protein profiling for the detection of breast cancer

M.E. de Noo, M.P.J. van der Werff, B.J. Mertens, H. Dalebout, M. Bladergroen, R.A.E.M. Tollenaar, A.M. Deelder

Submitted

ABSTRACT

Background

With over 1 million new cases in the world each year, breast cancer is the commonest malignancy in women and comprises 18% of all female cancers. Proteomic expression profiling generated by mass spectrometry has been suggested as a potential tool for the early diagnosis of cancer. The objective of our study was to assess and validate the feasibility of this approach for the detection of breast cancer.

Methods

In a randomised block design pre-operative serum samples obtained from 63 breast cancer patients and 73 controls were used to generate high-resolution MALDI-TOF protein profiles as a calibration set. The median age of the patient and control group was respectively, 52 (20-81) and 57 years (39-87). The MALDI-TOF spectra generated using WCX magnetic beads assisted mass spectrometry (Ultraflex) were smoothed, binned and normalised after baseline correction. After pre-processing of the spectra, linear discriminant analysis with double cross-validation, based on principal component analysis, was used to classify the protein profiles. Consequently, the classifier constructed on the first 2 plates was applied on the spectra of an independent validation set. This validation set consisted of serum samples from 29 breast cancer patients and 38 controls. The median age was 59 years (26-87) and 57 years (24-71) for the patient and control group respectively.

Results

Double cross-validatory analysis carried out on the protein spectra of the calibration set yielded a total recognition rate of 86%, a sensitivity of 88% and a specificity of 84% for the detection of breast cancer within the calibration set. The AUC of this classifier was 90.3%. When this classifier was applied on the spectra of the independent validation set a total recognition rate of 80.9%, a sensitivity of 72% and a specificity of 89% were found.

Conclusions

The use of a randomised block design, but mainly an independent validation set proves that discriminating protein profiles can be detected between breast cancer patients and healthy controls. Although further validation in larger series and identification of the discriminating proteins must be achieved, the high sensitivity and specificity indicate that serum protein profiles could be an option for the detection of breast cancer.

INTRODUCTION

With over 1 million new cases in the world each year, breast cancer is the commonest malignancy in women and comprises 18% of all female cancers. In 2005, breast cancer caused 502,000 deaths (7% of cancer deaths; almost 1% of all deaths) worldwide.[1] Despite increasing incidence rates, annual mortality rates from breast cancer have decreased over the last decade (2.3% per year from 1990 to 2002).[2] The effect of reduction due to early diagnosis of breast cancer has been outlined with patients' data by the Surveillance, Epidemiology, and End Results program in a competing-risk analysis calculating probabilities of death from breast cancer and other causes according tot stage, race and age at diagnosis.[3]

Currently, mammography remains the most important diagnostic tool in women with breast tissue that is not dense, although MRI and ultrasonography are used in case of impairment of the latter diagnostic results.[4] In many countries mammography is used as a population based screening in woman older than 50 years. The effect of breast screening in terms of breast cancer mortality reduction persists after long-term follow-up. A recent meta-analysis of seven randomised trials concluded that there was a 15-20% reduction in risk of death from breast cancer in women attending mammography.[5] However, up to 20% of new breast cancers are not detected or visible on a mammogram.[6]

Prognosis and selection of therapy may be influenced by the age and menopausal status of the patient, Bloom-Richardson stage, histological and nuclear grade of the primary tumour, oestrogen-receptor (ER) and progesterone-receptor (PR) status and HER2/neu gene amplification.[7] Currently, serum tumour markers play no role of importance in the diagnosis of breast cancer due to a lack of sensitivity and specificity.

Proteomic expression profiles generated with mass spectrometry have been suggested as a potential tool for the early diagnosis of cancer and other diseases. Although promising results have been shown in classifying cancer studies based on biomarker detection with multiple low-molecular-weight serum proteins[8-11] stringent demands have been proposed on both study design and experimental procedures for proteomic profiling.[12-14] Subsequently, several groups have stressed the importance of standardised protocols and homogeneity of subject groups and especially validation of the classification method.[15-19] Since no serum biomarker is currently available to detect breast cancer, the present study was designed to test and validate whether serum protein profiles generated with mass spectrometry could be indicative of the presence of breast cancer in an independent set.

MATERIAL AND METHODS

Subjects

Serum samples were obtained from a total of 111 patients one day prior to surgery for breast disease. All surgical specimens were histological examined and if malignant, the extent of tumour spread was assessed by TNM classification. Next to invasive stages of breast cancer, ductal carcinoma in situ (DCIS) samples were present in the patient group. The control group consisted of 92 healthy female volunteers. Patients and controls were included from October 2002 till July 2006 in our center.

In the calibration set the mean age of the patient group and control group was 52 (20-81) and 57 years (39-87) respectively. In the validation set the mean age was 59 years (26-87) and 57 years (24-71) for the patient and control group respectively (Table 1).

Table 1. Patient characteristics.

	Calibration set		Validation set	
	Patients	Controls	Patients	Controls
N=	73	63	38	29
Age (median)	52	57	57	59
(range)	20-81	39-87	26-87	24-71

Serum samples

Informed consent was obtained from all subjects and the study was approved by the Medical Ethical Committee of the LUMC. All samples were collected and processed following a standardised protocol: all blood samples were drawn from non-fasting patients or healthy controls while they were seated. The samples were collected in a 8.5 cc Serum Separator Vacutainer Tube (BD Diagnostics, Plymouth, UK) and centrifuged 30 min later at 3000 rpm for 10 minutes. The serum samples were distributed into 0.5 ml aliquots and stored at -70ºC. After thawing on ice the serum samples were randomised over different 96-well microtitration racks (Matrix, Hudson, USA) and then stored at -70ºC until the experiment.

Study design

We used a randomised blocked design to avoid any potential batch effects.[20;21] All the available samples from both groups were randomly distributed across 3 plates in roughly equal proportions (Table 2). For breast cancer, the distribution of disease stages across plates was again in random fashion and in approximately equal pro-

Table 2. Distribution and randomisation of serum samples of breast cancer patients and controls over the 3 MS target plates. Plate 1 and 2 were used as a calibration set, while plate 3 was used as a validation set.

	Plate 1	Plate 2	Plate 3	Total
Breast cancer	36	37	38	111
Controls	30	33	29	92
Total	66	70	67	203

Table 3. Distribution and randomisation of all different stages of breast cancer over the 3 MS target plates. Plate 1 and 2 were used as a calibration set, while plate 3 was used as a validation set.

Stage	Plate 1	Plate 2	Plate 3	Total
DCIS	3	7	7	17
I	12	14	14	40
IIA	11	11	6	28
IIB	5	3	7	15
IIIA	3	1	2	7
IIIB	1	1	2	4
IIIC	1	0	0	1
Total	36	37	38	111

portions (Table 3). The position on the plates of samples allocated to each plate was randomised as well. Each plate was then assigned to a distinct day. Analysis was carried out on 3 consecutive days, Tuesday to Thursday, processing a single plate each day. During the first two days, a calibration set with serum samples from 73 breast cancer patients and 63 controls was used to generate high-resolution MALDI-TOF protein profiles. The last day of the experiment, an independent validation set with serum samples from 38 breast cancer patients and 29 controls was measured.

Isolation of peptides and protein profiling

The isolation of peptides from serum was performed using magnetic beads based weak cation exchange chromatography (MB-WCX) kit from Bruker, mainly according to the manufacturers instructions, and adapted for automation on a 8-channel Hamilton STAR® pipetting robot (Hamilton, Martinsried, Germany). Magnetic beads with WCX-functionality (MB-WCX) were divided in 10 µl -aliquots in a 96-well microtiter plate, which was placed on the magnetic beads separation device (MPC®-auto96, Dynal, Oslo, Norway), with the magnet down. MB-WCX binding solution

(10 µl) and 5 µl serum sample were added to the beads and carefully mixed using the mixing feature of the robot. The sample was incubated for 5 minutes and the magnet was lifted, followed by a 30s waiting interval to settle the magnetic beads. The supernatant was removed and the magnet was lowered again. The magnetic beads were washed three times with MB-WCX washing solution (also provided with the kit) lifting and lowering the magnet as needed. The peptides were eluted from the beads using 10 µl elution solution (from the kit). Stabilization buffer was added (10 µl) and 2 µl of the stabilised eluate was transferred to a fresh 384-well microtiter plate (Greiner). Fifteen µl of α-cyano-4-hydroxycinnamic acid (0.3 g/l in ethanol: acetone 2:1) was added to the 2 µl eluate in the 384-well microtiter plate and mixed carefully. One microliter of this mixture was spotted in quadruplicate on a MALDI AnchorChip™ (Bruker Daltonics, Bremen, Germany).

Data processing and statistical analysis

To increase robustness, the average of four spots was used to represent one serum sample. All unprocessed spectra were exported from the Ultraflex in standard 8-bit binary ASCII format. They consisted of approximately 32,670 mass-to-charge ratio (m/z) values, covering a domain of 960 - 11,168 Dalton. The high-resolution spectra were first lightly smoothed and then, due to the quadratic nature of the TOF-equation, binned using a linear function of the time scale, resulting in bin widths of approximately 0.4 Dalton at the beginning of the spectrum and 1.4 Dalton at the end at the mass/charge scale. Subsequently, we normalised the spectra after baseline correction. In the calibration set, classification error rates were estimated and validated based on a classical Fisher linear discriminant analysis through complete double cross-validation as previously described.[24] This double cross-validated classifier was then applied on the validation set. This set was pre-processed using the exact same procedure as the calibration set. Using the estimated parameters from the calibration set, each sample in the validation set was assigned to the group for which the probability was highest. The error rates are based on sensitivity and specificity values, assuming a prior class probability of 0.5 for each group.

RESULTS

Three different randomised target plates were successfully measured on three consecutive days in the middle of the week. Figure 1 shows a raw data spectrum, directly obtained from the MALDI-TOF mass spectrometer. For further analysis, we first calculated the mean spectrum of each sample across all four spots that were measured for each sample, after pre-processing. The above-described pre-processing

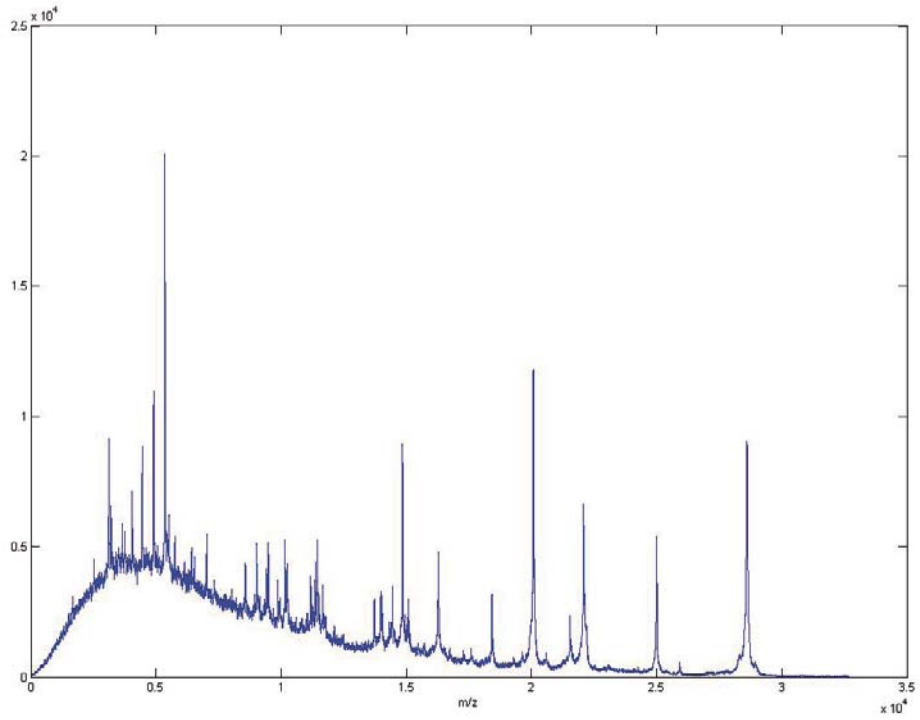


Figure 1. MALDI-TOF spectrum of a breast cancer patient after peptide isolation with WCX magnetic beads. On the Y-axis the relative intensity is shown. The mass to charge ration (m/z) is demonstrated on the X-axis in Dalton.

steps resulted in a sequence of 11,205 normalised m/z values ranging from 960 to 11,168 Dalton, for each individual.

Double cross-validatory analysis and evaluation carried out on the protein spectra of the calibration set (2 target plates) correctly classified 56 of 63 breast cancer patients as malignant. Sixty-one of 73 controls were correctly classified as non-cancer (Table 4a). The misclassified patients in the calibration set included 1 patient with DCIS, 4 stage I patients, 3 with stage IIA and 4 patients with stage IIB. There was no correlation with hormonal status.

These double cross validated results yielded a total recognition rate of 86%, a sensitivity of 88% and a specificity of 84% for the detection of breast cancer within the calibration set. To analyze the actual discriminative power of the classifier, we produced an ROC-curve (again based on the double cross-validatory classification probabilities), visualizing the performance of the two-class classifier in figure 2. The AUC of the classifier was 90.3%. To further evaluate possible bias of the double cross-validatory calculations, we performed a permutation exercise, which randomly permutes and reassigns the class labels across subjects and then repeats the entire

Table 4a. Double cross-validatory classification of serum samples **in calibration set**. A positive test results assigns subjects to the breast cancer (BC) group and a negative to the controls.

	Test results for detection of BC		
	Pos	Neg	Total
BC patients	61	12	73
Controls	7	56	63
	68	68	136

Table 4b. Double cross-validatory classification of serum samples **in validation set**.

	Test results for detection of BC		
	Pos	Neg	Total
BC patients	4	34	38
Controls	21	8	29
	25	42	67

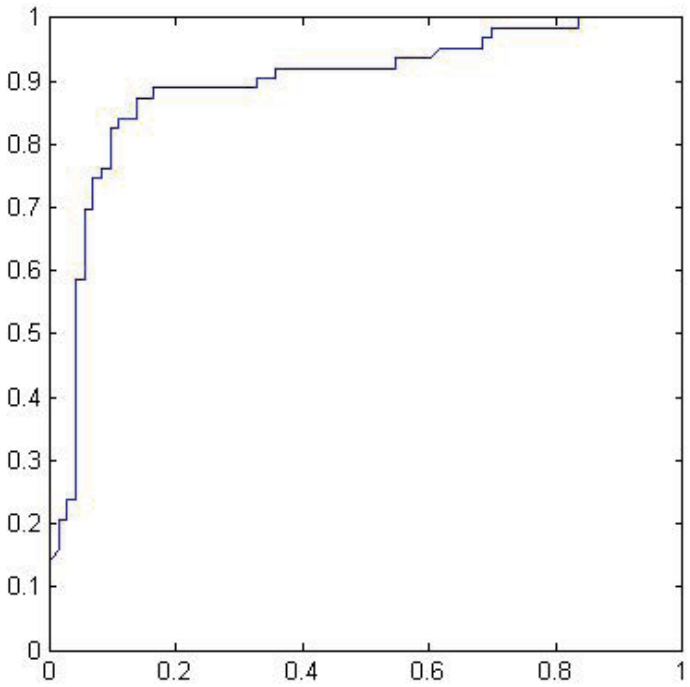


Figure 2. ROC-curve for the double cross-validated two-group classifier. The true positive recognition rate (sensitivity) is demonstrated on the y-axis against the false negative recognition rate (1-specificity) on the x-axis of the classifier.

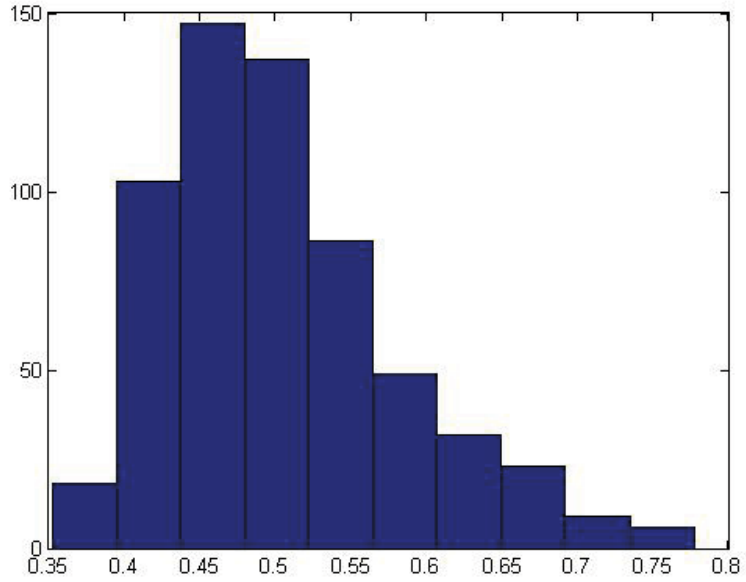


Figure 3. Histogram showing the normal distribution for the misclassification rate in the permutation exercise. The X-axis shows the misclassification rate calculated in the permutation exercise. On the Y-axis the number of permutations is displayed (n=600).

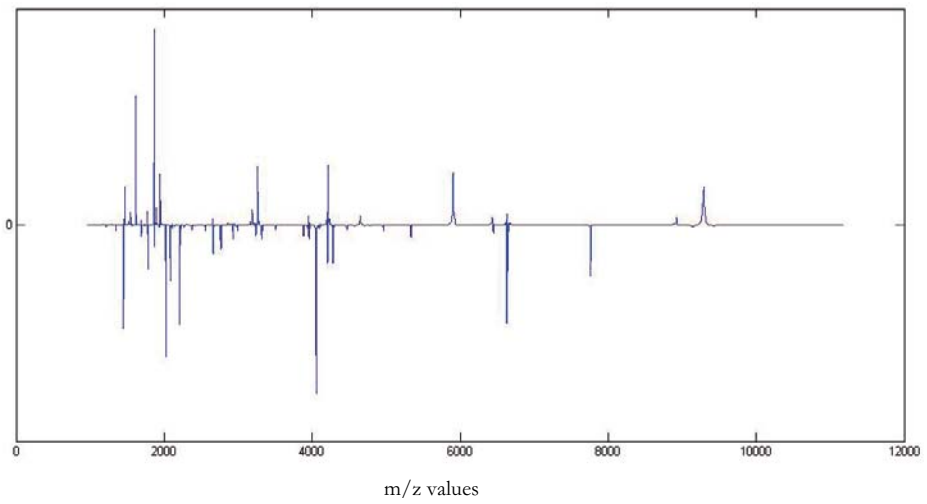


Figure 4. Correlation coefficients of most discriminating principal components with the class indicator. The correlation coefficients were calculated from the linear discriminant weightings.

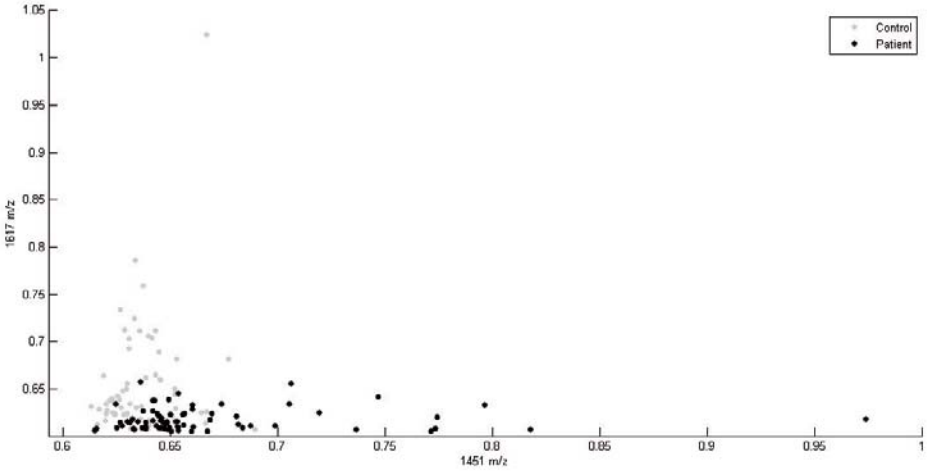


Figure 5a

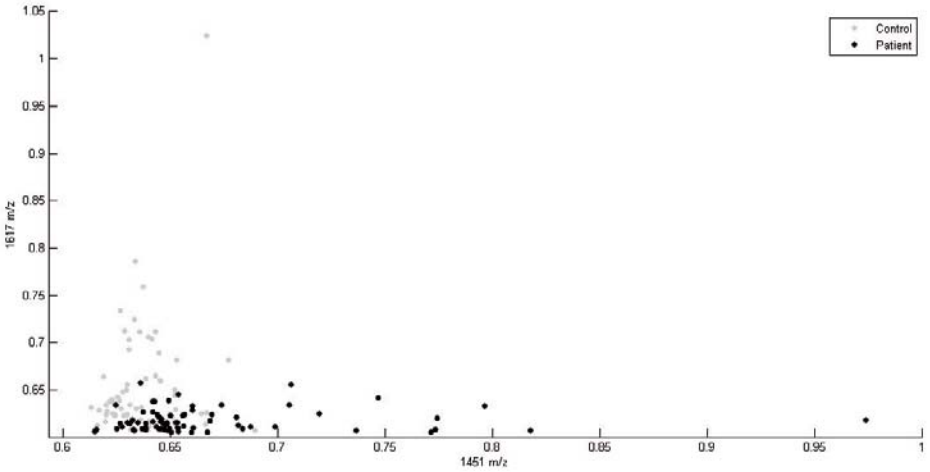


Figure 5b

Figure 5. Scatter plot of the first two principle components on basis of which the classification patient-control group was made.

double cross-validation procedure. Carrying out this procedure more than 600 times resulted in a median recognition rate of 49.0% with a 95% confidence interval of [0.39, 0.69] as shown in figure 3. The median AUC was 49.7% with confidence interval of [0.28, 0.61]. As both median recognition rates and AUC's equal roughly 50.0%, there is thus no substantial evidence of bias remaining within the cross-validatory calculation. Moreover, the actually observed recognition rates as well as AUC are

clearly separated from these permutation-based null-hypothesis confidence intervals, which prove the existence of discriminatory information in the spectra.

A post hoc exploration of the classification model was performed. In the present study 8 peaks that correlated most with the two groups provided most of the between-group separation. Figure 4 shows a plot of the correlation coefficients, with the class indicator, which can be calculated from the linear discriminant weightings in the region between 960 and 11,168 Dalton. Figure 5 shows scatter plots of 4 of these most discriminating peaks between cases from controls.

Consequently, we applied the double cross-validated classifier constructed on the first 2 plates on the spectra of the independent validation set. In this set 21 of 29 controls were correctly classified as non-malignant and only 3 of 38 breast cancer patients were misclassified, as shown in table 4b. From the 4 misclassified patients 1 had DCIS, whereas 2 patients both had stage I disease, according to their post-operative histological report. Nevertheless, these results produce a total recognition rate of 80.9%, a sensitivity of 72% and a specificity of 92% of the classifier in an independent dataset.

DISCUSSION

This validation study shows that breast cancer can be detected by serum protein profiling. We were able to classify breast cancer patients and healthy individuals accurately based upon information in MALDI-TOF serum spectra. In the calibration set the double cross-validated classifier demonstrated a sensitivity and specificity of 88% and 84% respectively. Sixty-one out of 73 controls were correctly classified as non-cancer. Moreover, with a lifetime risk of 1 in 9, it cannot be excluded that some of the control subjects currently are developing or carrying the disease. Unfortunately, since the control group consisted of anonymous symptom-free subjects it was impossible to retrieve the current physical state. The misclassified cancer patients had varying early stages of disease, from stage I to IIB. However, the fact that all but one DCIS patients was recognised in the cancer group, adds to its possible future applicability as a tool for early detection.

More importantly, in the independent validation set the classifier demonstrated a sensitivity and specificity of 72% and 92% respectively. While for the misclassified controls in the validation set the current physical state was also unknown due to their anonymity, the current physical state could easily be retrieved for the misclassified patients. Histological reports showed that one of them had DCIS and two of them were diagnosed with stage I disease. In this case the breast tumour had been resected for 2 years. Interestingly, this protein profile was classified in the

non-cancer group, which was confirmed from the treatment chart at the time of blood collection. When this sample would have been excluded of the analysis, the specificity of the classifier would increase even more in the validation set. Especially this high specificity adds to the potential of serum protein profiles to screen women at high risk for breast cancer, since there appears to be a low chance of false positive test results and unnecessary treatment will be avoided. When combined with mammography the positive predictive value of the proteomic pattern approach could be assessed in a high risk population.[22]

Since a potential drawback of any approach with high dimensional data is the tendency to discover patterns among many variables that may not be a direct result of the pathological state but rather a result of pre-analytical characteristics the need of an independent validation set has been stressed extensively.[14;23;24] In our previous studies, the use of an independent validation set was not possible due the relatively small sample. Therefore, until now we advocated a thorough and stringent study design and double cross-validation of the classification model.[24] This procedure avoided the need for separate test and validation sets to yield unbiased error rate estimates. However, in the current study discriminating protein profiles for the detection of breast cancer could be validated using an independent dataset. Nevertheless, the classifier in the calibration set was constructed following stringent demands. Again, a randomised block design was used to avoid observational bias, ensuring that no batch effects were introduced and artificial between-group separations excluded.[20;21] However, the issue clinically most relevant is the use of an independent validation set for the classification of diseased versus healthy individuals. This is primarily based on a specific problem in the discovery-based research field of clinical proteomics, namely overfitting. Overfitting may occur when multivariate models show apparent discrimination that is actually caused by data over-interpretation, and hence give rise to results that are not reproducible.[14;17;18] Therefore, protection against overfitting of the classifier was maintained by using the double cross validation in the calibration set. In this way, maximal reliability of the classifier was obtained by this procedure. Then, the performance of the classifier was tested in the independent validation set and it proved that breast cancer can indeed reliably be detected by discriminating protein profiles.

Obviously, other most common pitfalls in clinical proteomics such as sample collection, pre-analytical conditions and biological variation were avoided.[15;16;25] Therefore, serum sample collection and pre-analytical factors were rigorously standardised.[19] Furthermore, subjects in both groups were matched for age, although age is recently shown not to bias serum peptodomics.[26] In addition, patient samples from all stages of breast cancer were randomly distributed over three different target plates, excluding these factors as a discriminator in the current classifier.

Interestingly, the classification between cancer and non-cancer was performed using more principal components than in our earlier work when C8 magnetic beads were used. In the present study more than 2 peaks were responsible for most on the between group separation. As shown in figure 2 and 5, these include peaks with 1451, 1617, 5906 and 6644 m/z. Since the primary focus of this study was to assess and validate the feasibility of protein profiles based detection of breast cancer and therefore merely concentrated on pattern diagnostics, we have not identified these potential biomarkers yet. However, the controversy about the use of protein profiles as a pattern diagnostic without analysis of the diagnostic biomarkers still remains to be solved for its clinical application. Some have argued that low molecular weight proteins in serum, the serum peptidome, is nothing but aspecific biological trash and therefore does not yield any reliable biomarkers in the currently technically available mass range.[27-29] Nonetheless, recently it was postulated that although discriminating peptides do indeed belong to well known coagulation and complement pathways, their patterns or signatures do most certainly indicate the presence of cancer.[26] This study showed that most of the cancer-type specific biomarker fragments were generated in patient serum by enzymatic cleavage at previously known endoproteage cleavage sites after the blood sample was collected.[30] Villanueva et al. postulated that these cancer-specific low molecular weight proteins in the serum peptidome are an indirect snapshot of the enzyme activity in tumour cells. We support their hypothesis that discriminating serum protein profiles are a compilation of surrogate markers for the detection and classification of certain types of tumours.

In conclusion, the present study demonstrated that serum protein profiles generated with mass spectrometry could be indicative of the presence breast cancer. In order to obtain most realistic estimates of the discriminating power of serum protein profiles a classifier was constructed in a randomised block design. Maximal reliability in classification was achieved through double cross-validation of the classifier while maintaining protection against overfitting. Principally the potential of proteomic signatures from high dimensional mass spectrometry data as highly reliable diagnostic classifiers for the detection of breast cancer was actually confirmed in the independent validation set. The fact that high sensitivity and specificity could be maintained in the validation set is the first steps towards a new diagnostic approach in breast cancer.

REFERENCES

1. Jemal,A., Siegel,R., Ward,E., Murray,T., Xu,J., Smigal,C., and Thun,M.J. (2006) Cancer statistics, 2006. *CA Cancer J.Clin.*, 56, 106-130.
2. Edwards,B.K., Brown,M.L., Wingo,P.A., Howe,H.L., Ward,E., Ries,L.A., Schrag,D., Jamison,P.M., Jemal,A., Wu,X.C., Friedman,C., Harlan,L., Warren,J., Anderson,R.N., and Pickle,L.W. (2005) Annual report to the nation on the status of cancer, 1975-2002, featuring population-based trends in cancer treatment. *J.Natl.Cancer Inst.*, 97, 1407-1427.
3. Schairer,C., Mink,P.J., Carroll,L., and Devesa,S.S. (2004) Probabilities of death from breast cancer and other causes among female breast cancer patients. *J.Natl.Cancer Inst.*, 96, 1311-1321.
4. Veronesi,U., Boyle,P., Goldhirsch,A., Orecchia,R., and Viale,G. (2005) Breast cancer. *Lancet*, 365, 1727-1741.
5. Gotzsche,P.C. and Nielsen,M. (2006) Screening for breast cancer with mammography. *Cochrane.Database.Syst.Rev.*,CD001877.
6. Astley,S.M. (2004) Computer-based detection and prompting of mammographic abnormalities. *Br J Radiol*, 77, S194-S200.
7. Simpson,J.F., Gray,R., Dressler,L.G., Cobau,C.D., Falkson,C.I., Gilchrist,K.W., Pandya,K.J., Page,D.L., and Robert,N.J. (2000) Prognostic Value of Histologic Grade and Proliferative Activity in Axillary Node-Positive Breast Cancer: Results From the Eastern Cooperative Oncology Group Companion Study, EST 4189. *J Clin Oncol*, 18, 2059-2069.
8. Adam,B.L., Qu,Y., Davis,J.W., Ward,M.D., Clements,M.A., Cazares,L.H., Semmes,O.J., Schellhammer,P.F., Yasui,Y., Feng,Z., and Wright,G.L., Jr. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.*, 62, 3609-3614.
9. Petricoin,E.F., III, Ornstein,D.K., Paweletz,C.P., Ardekani,A., Hackett,P.S., Hitt,B.A., Velassco,A., Trucco,C., Wiegand,L., Wood,K., Simone,C.B., Levine,P.J., Linehan,W.M., Emmert-Buck,M.R., Steinberg,S.M., Kohn,E.C., and Liotta,L.A. (2002) Serum proteomic patterns for detection of prostate cancer. *J.Natl.Cancer Inst.*, 94, 1576-1578.
10. Rai,A.J., Zhang,Z., Rosenzweig,J., Shih,I., Pham,T., Fung,E.T., Sokoll,L.J., and Chan,D.W. (2002) Proteomic approaches to tumor marker discovery. *Arch.Pathol.Lab Med.*, 126, 1518-1526.
11. Yanagisawa,K., Shyr,Y., Xu,B.J., Massion,P.P., Larsen,P.H., White,B.C., Roberts,J.R., Edgerton,M., Gonzalez,A., Nadaf,S., Moore,J.H., Caprioli,R.M., and Carbone,D.P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet*, 362, 433-439.
12. Hu,J., Coombes,K.R., Morris,J.S., and Baggerly,K.A. (2005) The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Brief.Funct.Genomic.Proteomic.*, 3, 322-331.
13. Coombes,K.R., Morris,J.S., Hu,J., Edmonson,S.R., and Baggerly,K.A. (2005) Serum proteomics profiling-a young technology begins to mature. *Nat.Biotechnol.*, 23, 291-292.
14. Ransohoff,D.F. (2004) Rules of evidence for cancer molecular-marker discovery and validation. *Nat.Rev.Cancer*, 4, 309-314.
15. Boguski,M.S. and McIntosh,M.W. (2003) Biomedical informatics for proteomics. *Nature*, 422, 233-237.
16. Villanueva,J., Philip,J., Entenberg,D., Chaparro,C.A., Tanwar,M.K., Holland,E.C., and Tempst,P. (2004) Serum Peptide profiling by magnetic particle-assisted, automated sample processing and maldi-tof mass spectrometry. *Anal.Chem.*, 76, 1560-1570.
17. Diamandis,E.P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J.Natl.Cancer Inst.*, 96, 353-356.

18. Baggerly,K.A., Morris,J.S., and Coombes,K.R. (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics.*, 20, 777-785.
19. de Noo,M.E., Tollenaar,R.A.E.M., Ozalp,A., Kuppen,P.J.K., Bladergroen,M.R., and Deelder A.M. (2005) Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal.Chem.*, 77, 7232-7241.
20. Box,G.E.P., Hunter W.G., and Hunter J.S. (1978) *Statistics for experimenters*. John Wiley & Sons, Inc..
21. Cox D.R. and Reid N. (2000) *The theory of the design of experiments*. Chapmann/Hall CRC.
22. Espinosa,E., Redondo,A., Vara,J.A., Zamora,P., Casado,E., Cejas,P., and Baron,M.G. (2006) High-throughput techniques in breast cancer: A clinical perspective. *Eur.J Cancer*.
23. Somorjai,R.L., Dolenko,B., and Baumgartner,R. (2003) Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics.*, 19, 1484-1491.
24. de Noo,M.E., Mertens,B.J., Ozalp,A., Bladergroen,M.R., van der Werff,M.P., van de Velde,C.J., Deelder,A.M., and Tollenaar,R.A. (2006) Detection of colorectal cancer using MALDI-TOF serum protein profiling. *Eur.J Cancer*, 42, 1068-1076.
25. Baumann,S., Ceglarek,U., Fiedler,G.M., Lembcke,J., Leichtle,A., and Thiery,J. (2005) Standardized Approach to Proteome Profiling of Human Serum Based on Magnetic Bead Separation and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Clin. Chem.*, 51, 973-980.
26. Villanueva,J., Shaffer,D.R., Philip,J., Chaparro,C.A., Erdjument-Bromage,H., Olshen,A.B., Fleisher,M., Lilja,H., Brogi,E., Boyd,J., Sanchez-Carbayo,M., Holland,E.C., Cordon-Cardo,C., Scher,H.I., and Tempst,P. (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest*, 116, 271-284.
27. Diamandis,E.P. and van der Merwe,D.E. (2005) Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin Cancer Res.*, 11, 963-965.
28. Sorace,J.M. and Zhan,M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC.Bioinformatics.*, 4, 24.
29. Koomen,J.M., Li,D., Xiao,L.C., Liu,T.C., Coombes,K.R., Abbruzzese,J., and Kobayashi,R. (2005) Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J Proteome.Res.*, 4, 972-981.
30. Liotta,L.A. and Petricoin,E.F. (2006) Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest*, 116, 26-30.