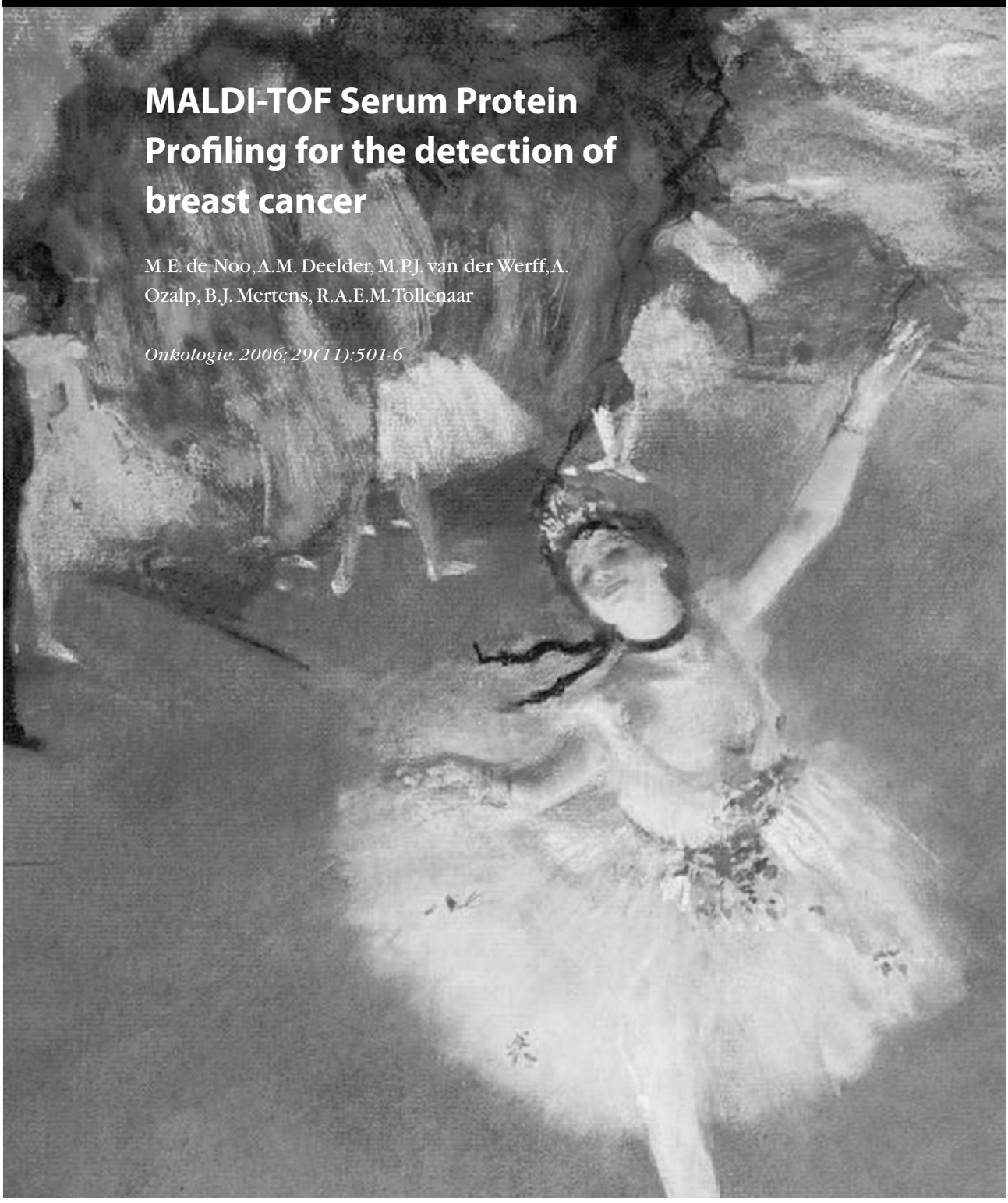
Clinical proteomics in oncology : a passionate dance between science and clinic

Noo, M.E. de

Chapter 6



MALDI-TOF Serum Protein Profiling for the detection of breast cancer

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

Onkologie. 2006; 29(11):501-6

ABSTRACT

Purpose

With a lifetime risk currently estimated one in nine, breast cancer is among the most common diagnosed malignancies and remains a leading cause of cancer-related morbidity and mortality. Proteomic expression profiling generated by mass spectrometry has been suggested as a potential tool for the early diagnosis of cancer and other diseases. The objective of our study was to assess the feasibility of this approach for the discrimination of breast cancer patients from healthy individuals.

Experimental design

In a randomised block design pre-operative serum samples obtained from 77 breast cancer patients and 29 controls were used to generate high-resolution MALDI-TOF protein profiles. The median age of the patient group and control group was respectively, 57.2 years and 50.0 years. All available 106 samples from both groups were randomly distributed across 3 plates in roughly equal proportions. The MALDI-TOF spectra generated using C8 magnetic beads assisted mass spectrometry (Ultraflex, Bruker Daltonics, Germany) 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.

Results

A total recognition rate of 99%, a sensitivity of 100% and a specificity of 97.0% for the detection of breast cancer were shown. The area under the curve of the classifier was 98.3%, which demonstrates the high, significant separation power of the classifier. The first 2 principal components account for most of the between-group separation.

Conclusions

Double cross-validation showed that classification could be attributed to actual information in the protein profiles rather than to chance. Although preliminary, the high sensitivity and specificity indicate the potential usefulness of serum protein profiles for the detection of breast cancer.

INTRODUCTION

With a lifetime risk currently estimated one in nine, breast cancer is among the most common diagnosed malignancies and remains a leading cause of cancer-related morbidity and mortality. Although the precise pathways of tumour genesis remain poorly defined, it appears that most invasive breast cancers arise from gene alterations that result in an initial transformation of normal breast tissue to in situ carcinoma.[1] Currently, mammography remains the most important diagnostic tool, although MRI and ultrasonography are used in case of impairment of the latter diagnostic results.[2] However, up to 20% of new breast cancers are not detected or visible on a mammogram.[3] 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, measures of proliferative capacity, and HER2/neu gene amplification.[4] 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 potential tools for the early diagnosis of cancer and other diseases. After the initial 'hype' of biomarker detection on the basis of multiple low-molecular-weight serum proteins stringent demands have been proposed on both study design and experimental procedures for proteomic profiling.[5-11] Subsequently, several studies appeared showing the importance of standardised protocols and homogeneity of subject groups and especially validation of the classification method.[12-16] This study aims to live up to all these demands.

Since no serum biomarker is currently known to reliably 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 breast cancer.

MATERIAL AND METHODS

Subjects

Serum samples were obtained from a total of 77 patients one day prior to surgery for a breast disease. All surgical specimens were histologically examined and if malignant, the extent of tumour spread was assessed by TNM classification. All stages of breast cancer were present in the patient group. The median age of the patient group was 57.2 years (range 32.6-90.3). Patients were included from October 2002 till July 2005 in our center. The control group consisted of 29 healthy female volunteers. The

median age of the healthy symptom-free control group was 50.0 years (25.9-76.7). The 29 controls were included in November and December 2004 (Table 1).

Table 1. Patient characteristics.

	Patients	Controls
n =	78	29
Age (mean)	56.6	49.9
(range)	36.2-90.3	25.9-76.7

Serum samples

Informed consent was obtained from all patients 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 while the patients or healthy controls were seated and non-fasting. The samples were collected in a 10 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 1 ml aliquots and stored at -70 °C. After thawing on ice the serum samples were randomised over different 96-well microtitration racks (Matrix) and then stored at -70°C until the experiment.

Study design

We used a randomised blocked design to avoid any potential batch effects.[17;18] All the available 106 samples from both groups were randomly distributed across 3 plates in roughly equal proportions (Table 2). For breast cancer, the distribution of stadia across plates was again in random fashion and in approximately equal proportions (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.

Table 2. Distribution and randomisation of serum samples of colorectal cancer patients with different TNM stage before and after the MALDI-TOF experiment. The distribution of stadia across plates was performed randomly random fashion and approximately equal proportions.

	Plate 1	**Plate 2**	**Plate 3**	**Total**
Breast cancer	26	26	26	78
Controls	11	9	9	29
Total	37	35	35	107

Table 3. Distribution and randomisation of serum samples of breast cancer group over the three MS target plates.

Stage	Plate 1	Plate 2	Plate 3	Total
DCIS	5	4	3	12
I	6	6	8	22
IIA	7	8	3	18
IIB	4	6	4	14
IIIA	1	2	4	5
IIIB	1	0	2	3
IIIC	1	0	2	3
Total	25	26	26	77

Isolation of peptides and protein profiling

The isolation of peptides from serum was performed using the C8 magnetic beads based hydrophobic interaction chromatography (MB-HIC) kit from Bruker Daltonics (Bremen, Germany) mainly according to manufacturers instructions, adapted for automation on a 8-channel Hamilton STAR® pipetting robot (Hamilton, Martinsried, Germany) as previously described by our group. Each sample was spotted in quadruplicate on a MALDI AnchorChip™. Matrix Assisted Laser Desorption Ionisation Time-Of-Flight (MALDI-TOF) mass spectrometry measurements were performed using an Ultraflex I TOF/TOF instrument (Bruker Daltonics, Bremen, Germany) equipped with a SCOUT ion source, operating in linear mode. Ions formed with a N2 pulse laser beam (337 nm) were accelerated to 25 kV. With this specific serum preparation peptide/protein peaks in the m/z range of 960 to 11,169 Dalton were measured.

Data processing and statistical analysis

All unprocessed spectra were exported from the Ultraflex in standard 8-bit binary ASCII format. They consisted of approximately 45,000 mass-to-charge ratio (m/z) values, covering a domain of 1.160 - 11,600 Dalton. To increase robustness, the average of four spots was used to represent one serum sample. Subsequently, we lightly smoothed, binned and normalised the spectra after baseline correction. Fully validated classification error rates were estimated based on a classical Fisher linear discriminant analysis through complete double cross-validatory joint estimation and assessment of class predictions as previously described.[19]

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. Before pre-processing and further analysis a mean spectrum of each sample was calculated over all four spots that were measured for each sample. The above-described pre-processing steps resulted in a sequence of 4483 normalised m/z values ranging from 1160 to 11,600 Dalton, for each individual. One sample from the breast cancer group was excluded from analysis due to its poor quality spectra.

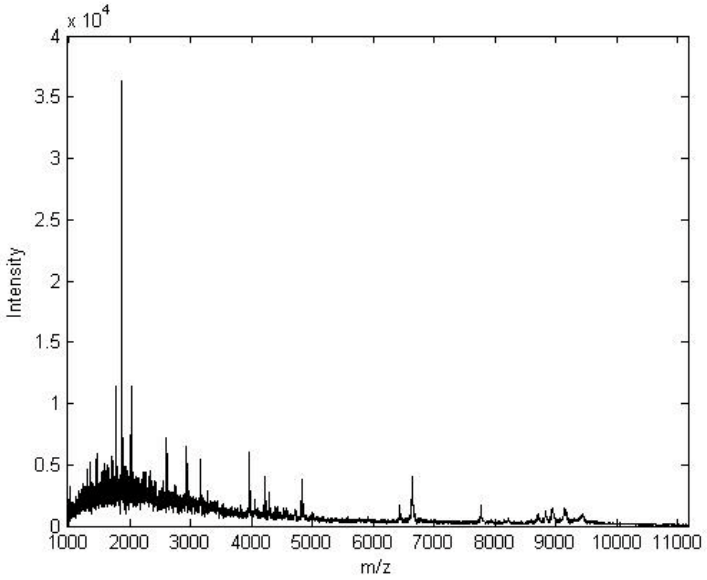


Figure 1. MALDI-TOF spectrum of a breast cancer patient after peptide isolation with C8 magnetic beads. On the Y-axis the relative intensity is shown. The mass to charge ration (m/z) is demonstrated on the X-axis in Dalton.

Double cross-validatory analysis and evaluation carried out on the protein spectra correctly classified 28 of 29 controls as non-cancer. All breast cancer patients were correctly classified as malignant (Table 4). These validated results yield a total recognition rate of 98.2%, a sensitivity of 100% and a specificity of 97.6% for the detection of breast cancer. 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 98.3%. The median AUC was 49.4% with confidence interval of

Table 4. Double cross-validatory classification of serum samples. A positive test results assigns subjects to the breast cancer (BC) group and a negative to the controls. In the horizontal plane the actual histologically confirmed diagnosis is stated.

	Test results for detection of BC		
	Neg	Pos	Total
Controls	77	0	77
CRC patients	1	28	29
	78	28	106

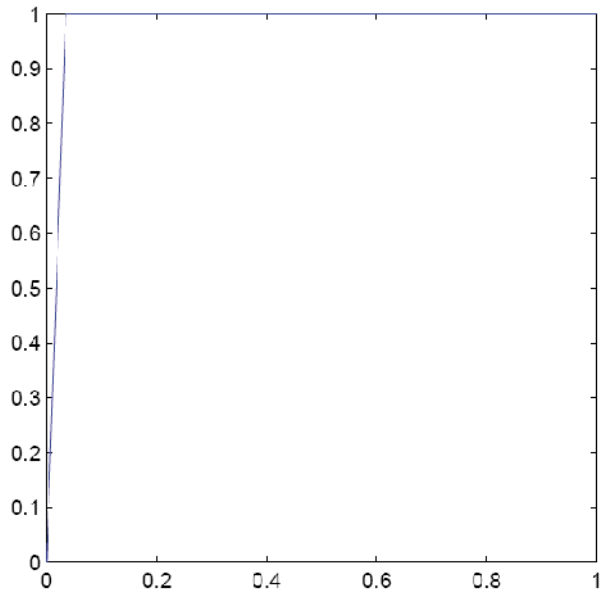


Figure 2. ROC-curve for the double cross-validated two-group classifier. The true positive recognition rate (sensitivity) is demonstrated on the y-axis against the false negative recognition rate (1-specificity) on the x-axis of the classifier.

[24.8, 64.2]. As both median recognition rates and AUC's equal 50%, there is thus no substantial evidence of bias remaining within the cross-validatory calculation.

We then proceeded to a post hoc exploration of the classification model. In the present study the first two principal components provided most of the between-group separation. Figure 3 shows a plot of the correlation coefficients, with the class indicator, which can be calculated from the linear discriminant weightings in the region between 1.160 and 11,600 Dalton.[20;21] As illustrated, the classification is achieved primarily through a contrast in peak intensities between the first and sec-

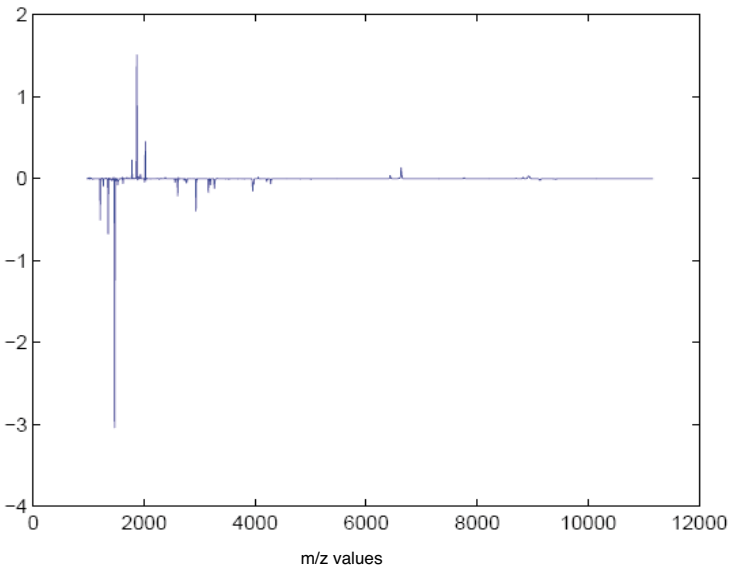


Figure 3. Correlation coefficients of two first principal components with the class indicator. The correlation coefficients were calculated from the linear discriminant weightings. The negative correlation of the first peak is an indicator for the control group and the positive correlation of the second peak points out the cases.

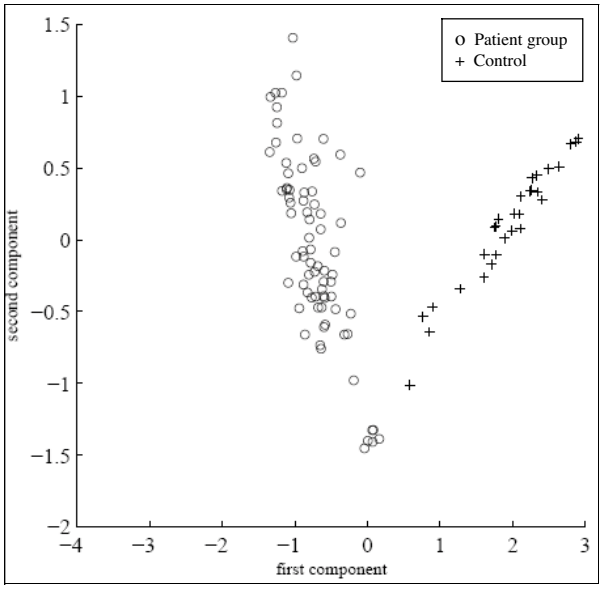


Figure 4. Scatter plot of the first two principle components on basis of which the classification patient-control group was made.

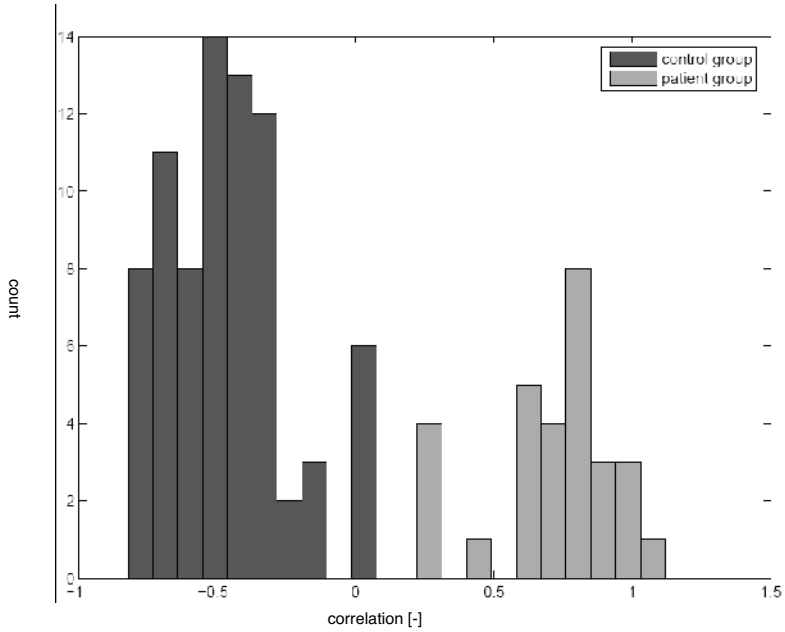


Figure 5. Histogram showing the difference between the normalized intensities of the two most discriminating "peaks" (bins). The X-axis shows the difference between the normalized intensities of the peaks. On the Y-axis the number of subjects is displayed.

ond principal component. This can also be seen in the scatter plot shown in figure 4: low intensities at the first peak for cases separates cases from controls. Likewise, a small contribution for controls at the second peak separates controls from patients. To illustrate these results further, we can simply calculate the contrast between the two peak intensities directly across all subjects and construct a simple one-dimensional summary of the data, as shown in the histogram displayed in figure 5, which shows overlapping histograms of this (ad hoc) contrast for each group separately. The separation is clearly visible. We also quantified the significance of this difference by performing a two-sample Student t-test on this contrast, ($p<0.0001$).

DISCUSSION

This study underlines the potential of serum protein profiling for the detection of breast cancer. We were able to classify breast cancer patients and healthy individuals very accurately based upon information in MALDI-TOF serum spectra. The classifier, calibrated and validated on spectra of the entire dataset demonstrated a sensitivity and specificity of 100% and 97.6% respectively. Only one subject from the control

group was misclassified in the malignant group (Table 4). Moreover, with a lifetime risk of one in nine, it might even be the case that one of 29 control subjects currently is 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. All patients with various stages of breast cancer were correctly classified, including DCIS patients. The fact that all DCIS patients are recognised in the cancer group, adds to its possible future applicability as a tool for early detection. If these results are validated, future studies could be performed to screen women at high risk for breast cancer by both mammography and serum protein profiling.[22] In that way the positive predictive value of the proteomic pattern approach could be assessed. Further, efforts have to be made to correlate different stages of breast cancer with serum protein profiles because this may contribute better prognostication and may eventually lead to more individualised treatment. Obviously, validation and sufficient sample size are once again of paramount importance for the reliability and its potential in a clinical setting.

We favour a thorough and stringent study design and double cross-validation of our classification model.[19] We feel that the use of standardised serum collection and mass spectrometry protocols, as advocated in various studies, has lifted serum protein profiling to a more reliable level.[12;13;16;23] To avoid the most common pitfalls in clinical proteomics sample collection, pre-analytical conditions and biological variables were in the present study matched for both groups and were rigorously standardised. The location of blood collection, i.e. the outdoor clinic for controls and the surgical ward for the patient group showed no influence on serum protein profiles (data not shown). Furthermore, 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. Ideally, the control group should consist of precisely age-matched individuals undergoing a mammography showing no aberrations. However, in practice this is difficult to realize, due to ethical and logistical issues. Notwithstanding, we performed an analysis to examine the differences in intensity of most discriminating peaks based on age. In the present study there was no significant contribution of one of these factors on the most discriminating peaks of our classification model (data not shown).

Regarding the bioinformatic and statistical approach of these high dimensional data there were two main points to consider: avoiding batch effects and validation of the classification model. To avoid observational bias, a randomised block design was used as an additional precaution. The randomised block design ensured that no batch effects were introduced and excluded artificial between-group separations. [17;18] Another recurrent topic of debate in serum protein profiling is validation of the classification model.[11;24] Consensus is achieved that, ideally, discriminating

protein profiles for the detection of a certain malignancy should be validated using an independent dataset. In the current phase of this study, the use of an independent validation set was excluded since the relatively small sample size did not allow this. Until a larger sample set is obtained, we advocate the use of a double cross validation of classification. This procedure avoids the need for separate test and validation sets to yield unbiased error rate estimates. The double validatory aspect of the procedure results from the fact that the discriminant rule constructed to classify the left-out data was optimised through a secondary cross-validatory evaluation within the first cross-validatory layer.[19;25]

The classification between cancer and non-cancer was mostly performed using the first two principal components, corresponding to two most discriminating peaks. Identification and functional analysis of these discriminating proteins/peptides might render new insights on tumour development and environmental responsiveness, which could eventually be translated in new diagnostic and prognostic insights for the clinician. Until nowadays, little success has been booked in assigning reproducible discriminating biomarkers.[14;24] Though this study showed two most discriminating mass values of MALDI-TOF based protein profiling analysis to be low molecular weight fragments, we have not identified these potential biomarkers yet. 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.[26-28] However, very recently Villanueva et al. published a study in which they proposed 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.[29] 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] They 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 to 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 patterns of proteomic signatures from high dimensional mass spectrometry data can be used as highly reliable diagnostic classifiers for the detection of breast cancer. With the double crossvalidatory study in a randomised block design we obtained maximal reliability in classification while maintaining protection against overfitting. Surely, independent validation and follow up studies are necessary and currently in progress. Nevertheless, the

extremely high sensitivity and specificity of the present study are highly promising for a new diagnostic approach in breast cancer.

REFERENCES

1. Burstein,H.J., Polyak,K., Wong,J.S., Lester,S.C., and Kaelin,C.M. (2004) Ductal carcinoma in situ of the breast. *N.Engl.J Med.*, 350, 1430-1441.
2. Veronesi,U., Boyle,P., Goldhirsch,A., Orecchia,R., and Viale,G. (2005) Breast cancer. *Lancet*, 365, 1727-1741.
3. Astley,S.M. (2004) Computer-based detection and prompting of mammographic abnormalities. *Br J Radiol*, 77, S194-S200.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Ransohoff,D.F. (2004) Rules of evidence for cancer molecular-marker discovery and validation. *Nat.Rev.Cancer*, 4, 309-314.
12. Boguski,M.S. and McIntosh,M.W. (2003) Biomedical informatics for proteomics. *Nature*, 422, 233-237.
13. 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.
14. 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.
15. 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.
16. 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.
17. Box,G.E.P., Hunter W.G., and Hunter J.S. (1978) *Statistics for experimenters*. John Wiley & Sons, Inc..
18. Cox D.R. and Reid N. (2000) *The theory of the design of experiments*. Chapmann/Hall CRC.

19. 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.
20. Seber,G.A.F. (2005) *Multivariate Observations*. John Wiley & Sons Inc.
21. Ripley,B.D. (2005) *Pattern recognition and neural networks*. Cambridge University Press.
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. 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.
24. 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.
25. Mertens,B.J.A. (2003) Microarrays, pattern recognition and exploratory data analysis. *Statistics in Medicine*, 22, 1879-1899.
26. 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.
27. Sorace,J.M. and Zhan,M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC.Bioinformatics.*, 4, 24.
28. 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.
29. 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.
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.