ORIGINAL PAPER

Postoperative serum proteomic profiles may predict recurrence-free survival in high-risk primary breast cancer

Marie-Christine W. Gast · Marc Zapatka · Harm van Tinteren ·
Marijke Bontenbal · Paul N. Span · Vivianne C. G. Tjan-Heijnen ·
Jaco C. Knol · Connie R. Jimenez · Jan H. M. Schellens · Jos H. Beijnen

Abstract

Purpose Better breast cancer prognostication may improve selection of patients for adjuvant therapy. We conducted a retrospective longitudinal study in which we investigated sera of high-risk primary breast cancer patients, to search for proteins predictive of recurrence-free survival.

Methods Sera of 82 breast cancer patients obtained after surgery, but prior to the administration of adjuvant therapy, were fractionated using anion-exchange chromatography, to facilitate the detection of the low-abundant serum peptides. Selected fractions were subsequently analysed by surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF MS), and the resulting protein profiles were searched for prognostic markers by appropriate bioinformatics tools.

Results Four peak clusters (i.e. m/z 3073, m/z 3274, m/z 4405 and m/z 7973) were found to bear significant prognostic value ($P \leq 0.01$). The m/z 3274 candidate marker was structurally identified as inter-alpha-trypsin inhibitor heavy chain 4 fragment$_{658-688}$ in serum. Except for the m/z 7973 peak cluster, these peaks remained independently

M.-C. W. Gast (✉) · J. H. Beijnen
Department of Pharmacy and Pharmacology, The Netherlands
Cancer Institute, Slotervaart Hospital, Amsterdam,
The Netherlands
e-mail: mcwgast@gmail.com

M. Zapatka
Intelligent Bio-Informatics Systems (iBIOS)-group,
German Cancer Research Center, Heidelberg, Germany

H. van Tinteren
Department of Biometrics, The Netherlands Cancer Institute,
Antoni van Leeuwenhoek Hospital, Amsterdam,
The Netherlands

M. Bontenbal
Department of Medical Oncology, Erasmus Medical
Center – Daniel den Hoed Cancer Center, Rotterdam,
The Netherlands

P. N. Span
Department of Laboratory Medicine, Radboud University
Medical Center, Nijmegen, The Netherlands

P. N. Span
Department of Radiation Oncology, Radboud University
Medical Center, Nijmegen, The Netherlands

V. C. G. Tjan-Heijnen
Division of Medical Oncology, Department of Internal
Medicine, Maastricht University Medical Centre,
Maastricht, The Netherlands

J. C. Knol · C. R. Jimenez
OncoProteomics Laboratory,
Department of Medical Oncology,
VU University Medical Center,
Amsterdam, The Netherlands

J. H. M. Schellens
Department of Medical Oncology,
Antoni van Leeuwenhoek Hospital,
The Netherlands Cancer Institute,
Amsterdam, The Netherlands

J. H. M. Schellens · J. H. Beijnen
Division of Biomedical Analysis,
Department of Pharmaceutical Sciences,
Faculty of Science, Utrecht University,
Utrecht, The Netherlands

associated with recurrence-free survival upon multivariate Cox regression analysis, including clinical parameters of known prognostic value in this study population.

Conclusion Investigation of the postoperative serum proteome by, e.g., anion-exchange fractionation followed by SELDI-TOF MS analysis is promising for the detection of novel prognostic factors. However, regarding the rather limited study population, validation of these results by analysis of independent study populations is warranted to assess the true clinical applicability of discovered prognostic markers. In addition, structural identification of the other markers will aid in elucidation of their role in breast cancer prognosis, as well as enable development of absolute quantitative assays.

Keywords Breast cancer · Recurrence-free survival · Prognostic markers · Serum · Anion-exchange fractionation · SELDI-TOF MS

Introduction

Breast cancer is at present the most commonly diagnosed neoplasm among women (Jemal et al. 2008). In addition, despite the substantial progress made in cancer therapy, breast cancer is the second leading cause of female cancer deaths, following lung cancer (Jemal et al. 2008). The main prognostic factors currently used to determine eligibility for administration of adjuvant systemic therapy include both clinical and pathological parameters, e.g., patient's age at diagnosis, tumour size, lymph node status, grade of malignancy (Goldhirsch et al. 2003). However, despite appropriate locoregional treatment and adjuvant systemic therapy, 30–50% of breast cancer patients will develop metastatic relapse and die (Early Breast Cancer Trialists' Collaborative Group 1998), while there is a substantial percentage of patients that would have survived without adjuvant chemo- and hormonal therapy. Evidently, currently applied prognostic markers do not suffice for precise risk-group determination in breast cancer. This failure most likely originates in the high molecular heterogeneity of breast cancer pathogenesis and progression, which the currently used prognostic parameters clearly cannot fully address. Improved prognostic markers that might help to reduce both over- and undertreatment of the disease are thus urgently needed.

In search for these markers, investigators from our institutes have published gene expression profiles in tumour tissue that outperformed all prognostic parameters in predicting disease outcome (i.e. distant metastases) ('t Veer et al. 2002; Foekens et al. 2006; van de Vijver et al. 2002; Wang et al. 2005). Nonetheless, it is currently understood that the functional "end-unit" of the genome, i.e., the proteome, might have greater ability in reflecting the molecular complexity of (breast) cancer. Covering posttranslational and posttranscriptional modifications, the proteome reflects both the intrinsic genetic programme of the cell and the impact of its immediate environment, providing a highly dynamic and accurate view of a biological status (Banks et al. 2000) and, hence, a rich and complementary source of potential biomarkers.

One of the proteomic technologies used extensively in the search for novel markers is surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF MS) (Hutchens and Yip 1993). By combining retention chromatography with laser desorption/ionisation MS instrumentation, this platform has enabled high-throughput mass profiling of highly complex biological samples, such as tissue lysates and serum. Thus far, only two studies have reported the use of SELDI-TOF MS for discovery of prognostic breast cancer markers (Goncalves et al. 2006; Ricolleau et al. 2006). Ricolleau et al. (2006) investigated tumour cytosolic extracts of 60 breast cancer patients and identified ubiquitin and ferritin light chain to be associated with prognosis. Goncalves et al. (2006), on the other hand, investigated serum, being an easier accessible biological matrix that provides a good reflection of the human proteome as it perfuses all tissues of the body. Following SELDI-TOF MS analysis of fractionated sera, they constructed a multiprotein model consisting of 40 proteins, correctly predicting relapse in 67 of 81 patients (Goncalves et al. 2006). Our research group has previously performed a prognostic SELDI-TOF MS study in serum as well (Gast et al. 2009). Although we initially discovered the haptoglobin phenotype to be a strong, independent, prognostic parameter in high-risk primary breast cancer ($n = 63$), this result most likely was false positive, as it was not confirmed following analysis of our validation sample set ($n = 371$) (Gast et al. 2009). In contrast to the study of Goncalves et al. (2006), we investigated raw, unfractionated sera in our previous study. While only 22 proteins comprise more than 99% of the human serum proteome, the low-abundant proteins make up for the remaining <1% (Anderson and Anderson 2002). This large dynamic range of proteins in crude serum hampers detection of the allegedly high-informative low-abundant serum proteins. Serum fractionation, however, is likely to facilitate detection of the low-abundant proteins through reduction of this dynamic range (Hoffman et al. 2007).

Hence, by analysis of fractionated serum, we aimed to obtain better sensitivity of detection of markers that can be applied in the prognostication of breast cancer. To this end, sera of 82 breast cancer patients procured after surgery, but prior to the administration of adjuvant therapy, were fractionated using anion-exchange chromatography. Selected

fractions were subsequently analysed by SELDI-TOF MS, and resulting protein profiles were searched for prognostic markers by appropriate bioinformatics tools.

Materials and methods

Study population

From 1993 to 1999, high-risk primary breast cancer patients who had undergone a modified radical mastectomy or breast conserving surgery with complete axillary clearance participated in a randomised, multicentre, phase III trial. This study investigated the benefit of high-dose adjuvant chemotherapy in patients with ≥4 axillary lymph node metastases. The design of the study has been described elsewhere (Rodenhuis et al. 2003). Major eligibility criteria were histologically confirmed stage 2A, 2B or 3A breast cancer with at least four tumour-positive axillary lymph nodes, but no evidence of distant metastases, age under 56 years, and no previous other malignancies.

In the current study, sera of 82 study patients who were treated in the Erasmus Medical Center–Daniel den Hoed Cancer Center (Erasmus: $n = 24$), or in the Radboud University Medical Center Nijmegen (Radboud: $n = 58$) were included. Sera were obtained after surgery (13–55 days), but prior to the administration of adjuvant chemotherapy (0–41 days), and all sera were stored at −80°C. All serum samples were obtained with medical ethics approval, and all patients gave informed consent.

Chemicals

All used chemicals were obtained from Sigma, St. Louis, MO, USA, unless stated otherwise.

Serum fractionation

Sera were fractionated manually using a strong anion-exchange Q ceramic resin (Bio-Rad Labs, Hercules, CA, USA), according to manufacturers' protocol. Briefly, sera (20 µl) were denatured in 9 M urea/2% 3[(3-cholamido-propyl)-dimethylammonio]-propane sulphonate (CHAPS), after which they were randomly allocated in duplicate to two 96-well ProteinChip Q filtration plates, prefilled with Q ceramic HyperD F resin (Bio-Rad Labs). In addition, one serum sample was randomly assigned to 12 different wells of each fractionation plate for quality control purposes. Following incubation (30 min), the flow through was collected using a vacuum manifold (Millipore, Billerica, MA, USA). Bound proteins were subsequently eluted with a stepwise pH gradient using wash buffers ranging from pH 9

to pH 3, followed by an organic buffer for elution of remaining proteins. As a result, six serum fractions (F) were obtained, i.e., F1 (flow-through plus pH 9), F2 (pH 7), F3 (pH 5), F4 (pH 4), F5 (pH 3) and F6 (organic buffer). Prior to protein profiling, fractions were stored overnight at +4°C.

SELDI-TOF MS protein profiling

Protein profiling of serum fractions was performed using the ProteinChip SELDI (PCS 4000) Reader (Bio-Rad Labs). Various array chemistries and fractions were initially evaluated to determine which combination provided the best protein profiles in terms of number and resolution of proteins. Following assay optimisation, we selected Immobilized Metal Affinity Capture (IMAC30) arrays for the analysis of F3 and F4 and weak cation-exchange (CM10) arrays for the analysis of F5 and F6. Throughout the manual assay, arrays were assembled in a 96-well bioprocessor, which was shaken on a MicroMix 5 platform shaker (DPC Cirrus Inc., Los Angeles, CA, USA) at setting 20/7.

IMAC30 arrays were charged with 50 µl of 100 mM copper sulphate (Merck, Darmstadt, Germany) for 10 min, followed by neutralisation (5 min) with 200 µl 100 mM sodium acetate buffer pH 4. Next, both IMAC30 and CM10 arrays were equilibrated twice for 5 min with 200 µl of their respective binding buffers (IMAC30: 0.01 M phosphate-buffered saline pH 7.4/0.5 M sodium chloride (Merck), CM10: 20 mM sodium acetate pH 4). Arrays were subsequently loaded with 85 µl of binding buffer and 15 µl of the fractionated sample. After incubation (30 min), arrays were washed three times with 200 µl of binding buffer, and following a quick rinse with MilliQ water (Millipore), arrays were air-dried. A 50% sinapinic acid (Bio-Rad Labs) solution in 50% acetonitrile (Lab-scan Ltd., Dublin, Ireland)/0.5% trifluoroacetic acid (Merck) was applied twice (1.0 µl) to the arrays as the matrix. Following air-drying, the arrays were analysed using the ProteinChip SELDI (PCS 4000) Reader. Data were collected between 0 and 300 kDa, averaging 530 laser shots with 3,500 nJ intensity, at focus mass 7.5 kDa and matrix attenuation 1,000 Da. For mass accuracy, the instrument was calibrated on the day of measurements with All-in-One protein standard (Bio-Rad Labs).

Statistics and bioinformatics

Mass spectrometry data were processed using the tbimass R-package (http://www.r-project.org, publication in preparation). After pre-processing (resampling, baseline correction, normalisation and alignment correction), peaks were recognised using PROcess (http://www.bioconductor.org)

on the mean spectra of each experimental group (fraction/ProteinChip array type). For the discovery of peak clusters with significant prognostic value, a subpopulation ($n = 68$) containing patients diagnosed with a recurrence within 36 months of follow-up ($n = 32$) and patients experiencing no recurrence after a follow-up of at least 48 months ($n = 36$) were extracted from the study population ($n = 82$). By investigating this subpopulation using Cox proportional hazards analysis, the peak clusters associated with recurrence were identified within all peaks of the combined data of all fractions/ProteinChip array types. For selection, a stepwise method was applied (i.e. stepBIC), an algorithm sequentially searching through all possible Cox proportional hazard models for the one that minimises the Bayesian Information Criterion (BIC). Recurrence-free survival was calculated from the date of randomisation to the time of first recurrence or death, or the date of last follow-up.

Clinical parameters were selected for known impact on recurrence in the study population, to prevent overfitting of the data by the model. To this end, a Cox proportional hazards analysis was performed including the known clinical parameters presented in Table 1, based on forward entry ($P < 0.05$). In addition, results were adjusted for treatment allocation as in the original clinical trial. The obtained model was subsequently applied to the study population.

A Cox proportional hazards model was subsequently build on the total study population, by inclusion of the relevant clinical parameters only. To investigate whether the relationship between peak intensities and recurrence-free survival could be explained by any of the relevant clinical parameters, the hazard ratios were adjusted for these clinical parameters by construction of a Cox proportional hazards model on the study population, incorporating the selected peak clusters and the relevant clinical parameters.

Since our study population originated from two different hospitals that allegedly used different sample collection protocols, our results could have been influenced by various pre-analytical factors. The influence of the different collection protocols on the SELDI-TOF MS protein profiles was investigated by multidimensional scaling of the SELDI-TOF MS spectra. Herewith, the degree of similarity or dissimilarity between the samples withdrawn at the two different hospitals is graphically expressed: points representing similarity tend to cluster together, while points representing dissimilarity tend to be far apart. The influence of collection centre on the protein profile was furthermore investigated by Cox proportional hazards analysis for each peak cluster separately, incorporating one peak cluster, relevant clinical parameters and collection centre.

The reproducibility of the assay was assessed by analysis of one quality control serum sample, fractionated 24 times by random assignment to 12 different wells of each of the two fractionation plates. Within the quality control spectra, all peaks with a signal-to-noise ratio (S/N) ≥ 2 were detected, after which the coefficient of variation was calculated on the corresponding peak intensities. Finally, for the candidate biomarkers, the per cent difference between replicates, divided by the mean peak intensity, was calculated for all replicates measured in the study population. All statistical tests were two sided, and $P < 0.05$ was considered statistically significant.

Peptide identification

For identification purposes, peptides of interest were extracted from serum (fractions) by reversed-phase C18 magnetic beads (Dynabeads RPC18, Invitrogen, Breda, The Netherlands) using a Kingfisher 96 liquid handling system (Thermo Fisher Scientific, Waltham, MA, USA), according to the optimized protocol described in (Jimenez et al. 2007). Briefly, sera were diluted in TFA 0.1%, after which the peptide content was bound to the beads. The beads were subsequently washed with 0.1% TFA and eluted with 50% ACN. Eluate (1 μl) was mixed with α-cyano-4-hydroxy-cinnamic acid matrix (2 μl), after which the mixture was spotted (0.7 μl) on a MALDI target plate. Analyses were performed on a 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Foster City, CA, USA). Fragment ion spectra were taken to search the NCBI 20081128 database (Homo sapiens: 216937 sequences) using the MASCOT search engine at http://www.matrixscience.com (Matrix Science Ltd., London, UK), with the following search parameters: monoisotopic precursor mass tolerance: 18 ppm, fragment mass tolerance: 1 Da, variable modifications: methionine oxidation, and no specified protease cleavage site.

Results

Study population

At time of analysis, 45 patients (Erasmus: 19 pts, Radboud: 26 pts) had a recurrence or had died and 37 patients (Erasmus: 5 pts, Radboud: 32 pts) were censored at a median follow-up of 6.5 years (Erasmus: 7.8 years, Radboud: 6.3 years). Patient characteristics are provided in Table 1. All patient characteristics were similarly distributed between the samples obtained from the Erasmus Medical Center and the Radboud University Medical Center, as determined by the chi-squared test or the Mann–Whitney U-test.

Table 1 Patient and tumour characteristics of the study population

	Erasmus (n = 24)		Radboud (n = 58)		Total (n = 82)	
	n	(%)	n	(%)	n	(%)
Patient characteristics						
Age, mean (range)	43.5 (26–54)		43.2 (28–54)		43.3 (26–54)	
<40 years	6	(25)	15	(26)	21	(26)
≥40 years	18	(75)	43	(74)	61	(74)
Menopausal status						
Premenopausal	21	(88)	52	(90)	73	(89)
Postmenopausal	3	(12)	5	(9)	8	(10)
Unknown	0	(0)	1	(1)	1	(1)
Surgery						
Mastectomy	15	(63)	40	(69)	55	(67)
Breast conserving	9	(37)	18	(31)	27	(33)
Treatment						
Conventional dose	13	(54)	30	(52)	43	(52)
High dose	11	(46)	28	(48)	39	(48)
Tumour characteristics						
Number of positive lymph nodes						
4–9	15	(63)	38	(66)	53	(65)
≥10	9	(37)	20	(34)	29	(35)
Tumour size						
T1 (<2 cm)	7	(29)	17	(29)	24	(29)
T2 (2–5 cm)	13	(54)	34	(59)	47	(57)
T3 (≥5 cm)	4	(17)	7	(12)	11	(14)
Her2/neu status						
Negative	13	(54)	37	(64)	50	(61)
Positive	10	(42)	19	(33)	29	(35)
Unknown	1	(4)	2	(3)	3	(4)
Oestrogen receptor status						
ER negative	9	(38)	23	(40)	32	(39)
ER positive	14	(58)	35	(60)	49	(60)
Unknown	1	(4)	0	(0)	1	(1)
Progesterone receptor status						
PR negative	10	(42)	24	(41)	34	(42)
PR positive	13	(54)	34	(59)	47	(57)
Unknown	1	(4)	0	(0)	1	(1)
Bloom-Richardson grade						
Grade I	0	(0)	7	(12)	7	(9)
Grade II	6	(25)	18	(31)	24	(29)
Grade III	17	(71)	32	(55)	49	(60)
Unknown	1	(4)	1	(2)	2	(2)

SELDI-TOF MS protein profiling

Out of the 400 peak clusters tested for inclusion into the model, four peak clusters [m/z 3073 (F4/IMAC30), m/z 3274 (F4/IMAC30), m/z 4404 (F6/CM10), m/z 7973 (F5/CM10)] were selected due to their significant association with recurrence-free survival in the subpopulation

(Table 2). A representative example of SELDI-TOF MS protein profiles is depicted in Fig. 1.

The parameters 'age' (<40 years, $P = 0.021$), 'number of positive lymph nodes' (≥10, $P = 0.012$), and 'progesterone receptor status' (negative, $P = 0.003$) showed to be significantly associated with reduced recurrence-free survival in the study population. In addition, the parameter

Table 2 Multivariate proportional hazards analyses for the risk of recurrence on selected peak clusters, before (model 1, subpopulation) and after (model 3, total study population) adjustment for relevant clinical parameters, and on relevant clinical parameters solely (model 2, total study population)

Parameter	Model 1—peak clusters			Model 2—clinical parameters			Model 3—combined		
	HR	(95% CI)	*P* value	HR	(95% CI)	*P* value	HR	(95% CI)	*P* value
Peak cluster									
m/z 3073	3.17	(2.03–4.96)	<0.001				2.48	(1.78–3.48)	<0.001
m/z 3274	10.18	(1.99–52.01)	0.005				11.71	(2.05–66.90)	0.006
m/z 4405	0.02	(0.01–0.25)	0.003				0.01	(0.01–0.17)	0.001
m/z 7973	0.05	(0.01–0.48)	0.010				0.24	(0.03–1.78)	0.160
Treatment									
CONV				1	–	–	1	–	–
HD				1.59	(0.86–2.95)	0.140	2.48	(1.24–4.98)	0.011
Age									
<40 years				1	–	–	1	–	–
≥40 years				0.44	(0.22–0.88)	0.021	0.35	(0.16–0.74)	0.006
No. of LN+									
≥10				1	–	–	1	–	–
4–9				0.44	(0.23–0.84)	0.012	0.37	(0.19–0.72)	0.003
PR status									
PR(−)				1	–	–	1	–	–
PR(+)				0.40	(0.22–0.73)	0.003	0.28	(0.14–0.55)	< 0.001

CONV conventional dose arm, *HD* high-dose arm, *LN+* number of positive lymph nodes, *PR* progesterone receptor status positive (+) and negative (−)

'treatment' was included to correct for the different treatment arms of the original clinical trial. Three of the four selected peak clusters (i.e. m/z 3072, m/z 3274 and m/z 4404) remained significantly associated with recurrence-free survival in combination with the clinical variables (Table 2).

Furthermore, using multidimensional scaling, we investigated the association of the different collection protocols (allegedly used by the two hospitals) on the SELDI-TOF MS serum protein profiles. As depicted in Fig. 2 for F4/CM10, spectra of the sera collected in the Erasmus Medical Center and the Radboud University Medical Center are randomly distributed, indicating no structural differences in the SELDI-TOF MS serum protein profiles of both hospitals. In addition, following Cox proportional hazards analysis including one peak cluster, relevant clinical parameters and collection centre, all peak clusters except m/z 7973, remained (borderline) significant (i.e. m/z 3073: HR = 3.44, $P = 0.046$, m/z 3274: HR = 2.39, $P = 0.051$, m/z 4405: HR = 0.107, $P < 0.001$, and m/z 7973: HR = 0.35, $P = 0.160$).

The reproducibility of the assay was investigated by calculation of the coefficient of variation of all peak clusters with S/N >2 detected in the quality control spectra ($n = 24$ per fraction/ProteinChip array type)

(Fig. 3). The median coefficient of variation of the peak intensities following fractionation and SELDI-TOF MS analysis ranged from 13.4 to 24.2% for the different fractions/ProteinChip arrays investigated, with an overall average CV of 20.2%. Of the four candidate markers measured in the study population, the median per cent difference in peak intensity between the replicates, divided by the mean peak intensity, ranged from 20 to 25% (Fig. 4).

Peptide identification

The MALDI serum(fraction) peptide profiles obtained after C18 magnetic bead-assisted peptide captures were searched for the presence of prognostic SELDI peaks based on mass matching. Due to the different chemistries used for peptide capture for SELDI-TOF MS (IMAC30 Cu) and MALDI-TOF MS (C18) and to the mass limitations for direct fragmentation, we were able to elucidate the identity of one of the four candidate prognostic peak clusters in the spectra of whole serum. The SELDI-TOF MS peak cluster at m/z 3274 was detected by MALDI-TOF/TOF MS as MH^+ ions at m/z 3271.69 (default calibration) and identified by MALDI-TOF/TOF MS/MS (Fig. 5) in conjunction with database searching as a

Fig. 1 Representative example of fractionated serum protein profiles of two patients (pt A: patient with no recurrence for 123 months, pt B: patient with a recurrence at 54 months)
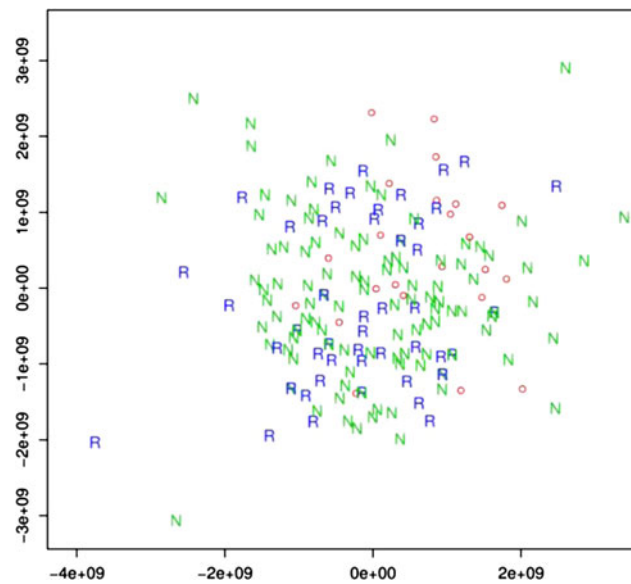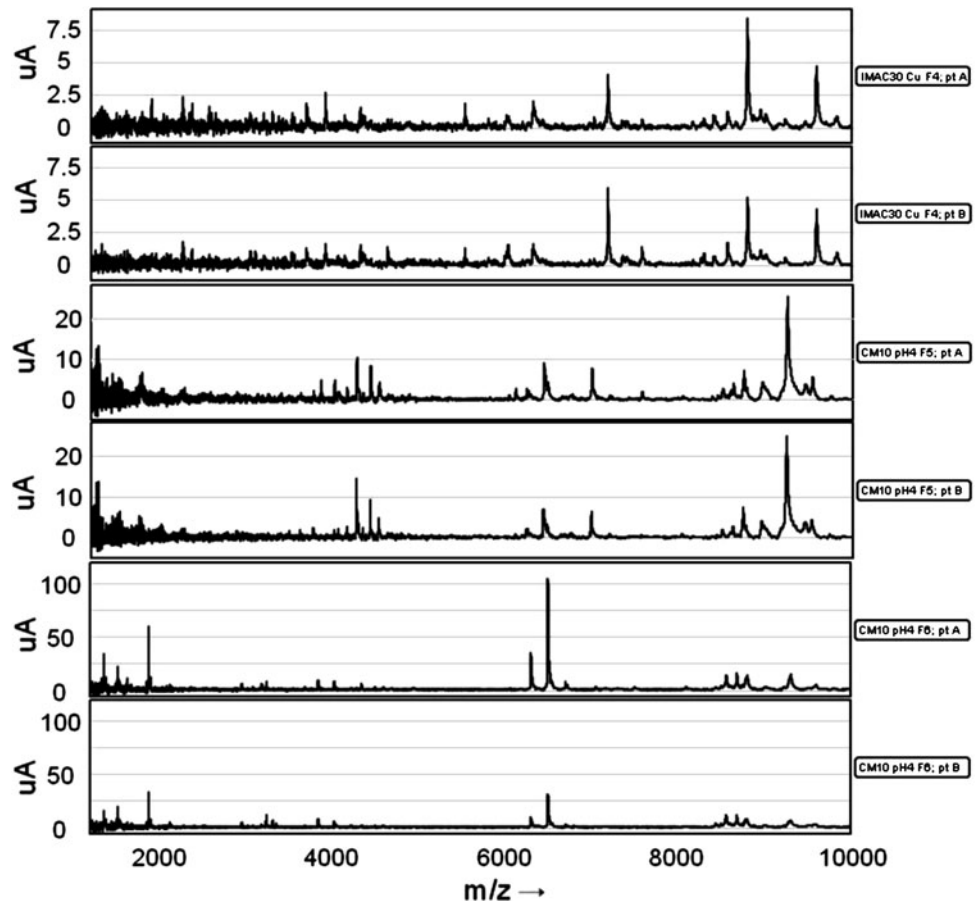

Fig. 2 MDS plot of Fraction 4/IMAC30 data (i.e. duplicate spectra) on center of withdrawal (*R* Erasmus, *N* Radboud, *O* quality control sample)

fragment of inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4$_{658-688}$), with a MASCOT score of 69 (expect: 0.0025).

Discussion

In the current study, we investigated sera of 82 breast cancer patients obtained after surgery, but prior to the administration of adjuvant therapy, in search for novel prognostic biomarkers. To facilitate the detection of the low-abundant serum peptides, sera were fractionated using anion-exchange chromatography, after which selected fractions were analysed by SELDI-TOF MS. Resulting protein profiles were searched for prognostic markers by appropriate bioinformatics tools. Considering solely the peak clusters detected in the SELDI-TOF MS protein profiles, four peak clusters (i.e. m/z 3073, m/z 3274, m/z 4405 and m/z 7973) were found to bear significant prognostic value, also after adjustment for known clinical prognostic parameters. The m/z 3274 candidate marker was structurally identified as ITIH4$_{658-688}$ in serum. Hence, investigation of the postoperative serum proteome by, e.g., anion-exchange fractionation, followed by SELDI-TOF MS analysis, is promising for the detection of novel prognostic factors. However, regarding the rather limited study population, validation of our results by analysis of similar, prospectively collected, independent, study populations is warranted to assess the true clinical applicability

Fig. 3 Coefficient of variation (*y*-axis) of the peak cluster identified in the quality control sample (*x*-axis) fractionated on fractionation plate 1 (*red*) and plate 2 (*green*)
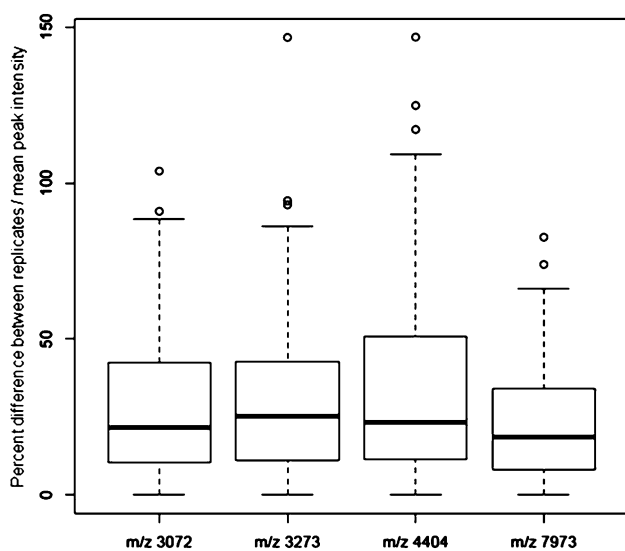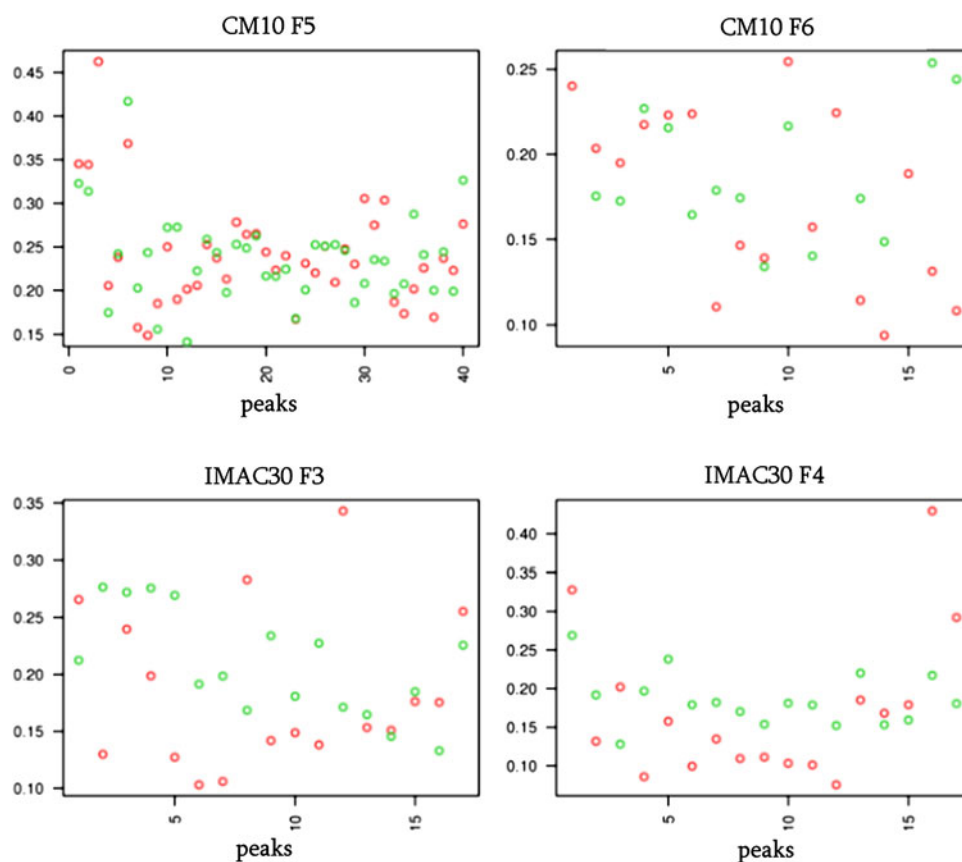

Fig. 4 Percent difference in peak intensity between replicates, divided by mean intensity, of the four candidate markers

of identified prognostic markers. In addition, structural identification of the other markers will aid in elucidation of their role in breast cancer prognosis, as well as enable development of absolute quantitative assays (e.g. van den Broek et al. 2008).

Metastases are thought to arise from clinically undetectable residual or micrometastatic disease, activated by a.o. stroma-generated growth factors, early impediment of immune surveillance and enhancement of angiogenesis (Demicheli et al. 1997, 2008; Heimann and Hellman 2000; Pupa et al. 2002). These early postsurgical host response processes are potentially affected by surgical extirpation of the tumour, as this disrupts the intricate interactions between malignant cells and physiological tumour-control mechanisms (Fisher et al. 1989; Tagliabue et al. 2003). Hence, the early postoperative serum proteome can bear prognostic information, since it reflects the host response processes that can play a key role in metastatic progression. The candidate prognostic markers detected in the current study therefore most likely correlate with this postoperative host response. In addition, since all study participants were treated with adjuvant chemotherapy, these differentially expressed proteins may also relate to the tumour phenotype and its chemosensitivity. Nonetheless, the four candidate markers could also arise directly from residual or micrometastatic disease. Considering the nadir in tumour burden following surgery, however, serum concentrations of tumour-secreted proteins most likely are well below the detection limit of the SELDI-TOF MS platform, even following serum fractionation. Lastly, the four candidate prognostic markers can also result from tumour-secreted

Fig. 5 Annotated MALDI-TOF/TOF MS/MS spectrum of m/z 3271.69
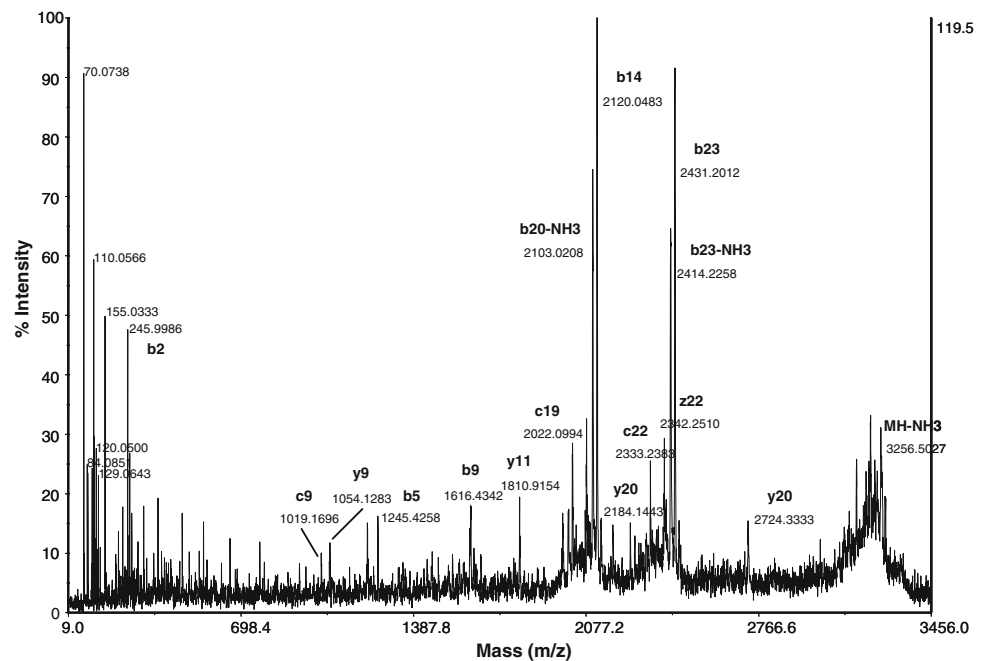
proteases that process host-response proteins upon their exposure to the tumour microenvironment (Fung et al. 2005; Villanueva et al. 2006). Since these modified host response proteins are generally present at substantially higher circulatory concentrations than the enzymes that process them upon their exposure to the tumour microenvironment, they can be detected in blood by SELDI-TOF MS.

This latter hypothesis is in fact endorsed by the structural identity of the candidate m/z 3274 marker, i.e., the inter-alpha-trypsin inhibitor heavy chain $4_{658-688}$ fragment, identified in serum. We previously found serum levels of this fragment decreased in breast cancer compared to control. Other studies have detected this fragment in serum as well, reporting either a lack of discriminative value (Fung et al. 2005; Villanueva et al. 2006) or an increase in breast cancer compared to control (Song et al. 2006). Most likely, these contradictory findings originate from the heterogeneity of the different study populations investigated or from the postulated instability of ITIH4 fragments (Fung et al. 2005; Song et al. 2006; Timms et al. 2007). In addition, changes in the abundance of the m/z 3274 ITIH4 fragment have been found associated with various types of cancer (e.g. prostate, breast, ovarian, colorectal and pancreatic cancer) (Fung et al. 2005; Villanueva et al. 2006; Song et al. 2006). This evident lack of specificity does not hamper its use as prognostic marker, however. The various serum ITIH4 fragments are currently hypothesised to result from tumour-secreted proteases that process host response proteins upon their exposure to the tumour microenvironment (Fung et al. 2005; Song et al. 2006; Villanueva et al.

2005). Hence, in the current study, the m/z 3274 marker could well originate from proteolytic activity associated with residual (micrometastatic) disease. According to this hypothesis, this candidate prognostic marker may be of value in other malignancies as well, as the protease activity has been shown to be cancer-type specific (Fung et al. 2005; Villanueva et al. 2006; Song et al. 2006). Hence, future validation studies should also include other types of malignancies.

Structural identification is imperative to investigate origin and function of the other three candidate biomarkers. In addition, concerning the rather limited study population, results must be validated by analysis of an independent, similar, sample set. However, such validation sets may prove difficult to obtain regarding the extended follow-up window needed to reliably investigate breast cancer prognosis.

While serum is generated by coagulation, its proteome is prone to the proteases involved in this cascade, as well as to those involved in the complement cascade, activated upon clotting. Various pre-analytical parameters, such as sampling device, clotting temperature and storage time, can thus all exert a distinct influence on the serum proteome. Since our study populations originated from two different hospitals that allegedly used different sample collection protocols, our results could have been influenced by the various pre-analytical factors. However, as depicted in Fig. 2, we did not observe such an influence on the protein profiles, indicating that the investigated serum proteome most likely is rather robust to (small) differences in collection protocols. Moreover, despite the different

characteristics of the two study groups, all peak clusters except m/z 7973 remained (borderline) significant after inclusion of the collection centre in the Cox proportional hazards model. The three peaks are therefore of additional prognostic value, even if the different collection centres are taken into account. The reliability of our results is furthermore endorsed by the reproducibility of the assay (average CV: 20.2%), which is well in agreement with previous reports (Goncalves et al. 2006; Albrethsen 2007).

Conclusion

In conclusion, using serum anion-exchange fractionation in combination with SELDI-TOF MS analysis, we discovered 4 peak clusters, one of which identified as serum ITIH4$_{658-688}$, with significant prognostic value in a study population of 82 high-risk primary breast cancer patients. Three peak clusters (including ITIH4$_{658-688}$) remained significantly associated with recurrence-free survival after adjustment for clinical parameters. These results are promising, as the prognostic profile identified in the current study could eventually improve patient selection. However, external validation using quantitative assays and elucidation of biological mechanisms are warranted.

Conflict of interest We declare that we have no conflict of interest.

References

Albrethsen J (2007) Reproducibility in protein profiling by MALDI-TOF mass spectrometry. Clin Chem 53(5):852–858

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1(11):845–867

Banks RE, Dunn MJ, Hochstrasser DF, Sanchez JC, Blackstock W, Pappin DJ et al (2000) Proteomics: new perspectives, new biomedical opportunities. Lancet 356(9243):1749–1756

Demicheli R, Retsky MW, Swartzendruber DE, Bonadonna G (1997) Proposal for a new model of breast cancer metastatic development. Ann Oncol 8(11):1075–1080

Demicheli R, Retsky MW, Hrushesky WJ, Baum M, Gukas ID (2008) The effects of surgery on tumor growth: a century of investigations. Ann Oncol 19(11):1821–1828

Early Breast Cancer Trialists' Collaborative Group (1998) Polychemotherapy for early breast cancer: an overview of the randomised trials. Lancet 352(9132):930–942

Fisher B, Gunduz N, Coyle J, Rudock C, Saffer E (1989) Presence of a growth-stimulating factor in serum following primary tumor removal in mice. Cancer Res 49(8):1996–2001

Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A et al (2006) Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 24(11):1665–1671

Fung ET, Yip TT, Lomas L, Wang Z, Yip C, Meng XY et al (2005) Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. Int J Cancer 115(5):783–789

Gast MC, van Tinteren H, Bontenbal M, van Hoesel QC, Nooij MA, Rodenhuis S et al (2009) Haptoglobin phenotype is not a predictor of recurrence free survival in high-risk primary breast cancer patients. BMC Cancer 8:389

Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ (2003) Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. J Clin Oncol 21(17):3357–3365

Goncalves A, Esterni B, Bertucci F, Sauvan R, Chabannon C, Cubizolles M et al (2006) Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene 25(7):981–989

Heimann R, Hellman S (2000) Individual characterisation of the metastatic capacity of human breast carcinoma. Eur J Cancer 36(13 Spec No):1631–1639

Hoffman SA, Joo WA, Echan LA, Speicher DW (2007) Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. J Chromatogr B Analyt Technol Biomed Life Sci 849(1–2):43–52

Hutchens TW, Yip TT (1993) New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Commun Mass Spectrom 7:576–580

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T et al (2008) Cancer statistics, 2008. CA Cancer J Clin 58(2):71–96

Jimenez CR, El Filali Z, Knol JC, Hoekman K, Kruyt FAE, Giaccone G et al (2007) Automated serum peptide profiling using novel magnetic C18 beads off-line coupled to MALDI-TOF-MS. Proteomics Clin Appl 1(6):598–604

Pupa SM, Menard S, Forti S, Tagliabue E (2002) New insights into the role of extracellular matrix during tumor onset and progression. J Cell Physiol 192(3):259–267

Ricolleau G, Charbonnel C, Lode L, Loussouarn D, Joalland MP, Bogumil R et al (2006) Surface-enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. Proteomics 6(6):1963–1975

Rodenhuis S, Bontenbal M, Beex LV, Wagstaff J, Richel DJ, Nooij MA et al (2003) High-dose chemotherapy with hematopoietic stem-cell rescue for high-risk breast cancer. N Engl J Med 349(1):7–16

Song J, Patel M, Rosenzweig CN, Chan-Li Y, Sokoll LJ, Fung ET et al (2006) Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. Clin Chem 52(6):1045–1053

't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536

Tagliabue E, Agresti R, Carcangiu ML, Ghirelli C, Morelli D, Campiglio M et al (2003) Role of HER2 in wound-induced breast carcinoma proliferation. Lancet 362(9383):527–533

Timms JF, Arslan-Low E, Gentry-Maharaj A, Luo Z, T'Jampens D, Podust VN et al (2007) Preanalytic influence of sample handling on SELDI-TOF serum protein profiles. Clin Chem 53(4):645–656

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25):1999–2009

van den Broek I, Sparidans RW, Schellens JH, Beijnen JH (2008) Liquid chromatography/tandem mass spectrometric method for the quantification of eight proteolytic fragments of ITIH4 with biomarker potential in human plasma and serum. Rapid Commun Mass Spectrom 22(18):2915–2928

Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L et al (2005) Correcting common errors in identifying cancer-specific serum peptide signatures. J Proteome Res 4(4):1060–1072

Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB et al (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 116(1):271–284

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F et al (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365(9460):671–679