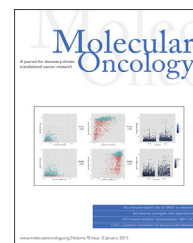


available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study



Leonardo Tenori^{a,b,1}, Catherine Oakman^{c,1}, Patrick G. Morris^d,
Ewa Gralka^{a,b}, Natalie Turner^c, Silvia Cappadona^c, Monica Fornier^d,
Cliff Hudis^d, Larry Norton^d, Claudio Luchinat^{a,e}, Angelo Di Leo^{c,*}

^aMagnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

^bFiorGen Foundation, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

^c‘Sandro Pitigliani’ Medical Oncology Department, Hospital of Prato, Via Suor Niccolina 20, Istituto Toscano Tumori, 59100 Prato, Italy

^dBreast Cancer Medicine Service, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

^eDepartment of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

ARTICLE INFO

Article history:

Received 18 March 2014

Received in revised form

14 July 2014

Accepted 15 July 2014

Available online 10 August 2014

Keywords:

Breast cancer

Biomarker

Metabolites

Metabolomics

Micrometastases

Nuclear magnetic resonance spectroscopy

ABSTRACT

Purpose: Metabolomics is a global study of metabolites in biological samples. In this study we explored whether serum metabolomic spectra could distinguish between early and metastatic breast cancer patients and predict disease relapse.

Methods: Serum samples were analysed from women with metastatic ($n = 95$) and predominantly oestrogen receptor (ER) negative early stage ($n = 80$) breast cancer using high resolution nuclear magnetic resonance spectroscopy. Multivariate statistics and a Random Forest classifier were used to create a prognostic model for disease relapse in early patients.

Results: In the early breast cancer training set ($n = 40$), metabolomics correctly distinguished between early and metastatic disease in 83.7% of cases. A prognostic risk model predicted relapse with 90% sensitivity (95% CI 74.9–94.8%), 67% specificity (95% CI 63.0–73.4%) and 73% predictive accuracy (95% CI 70.6–74.8%). These results were reproduced in an independent early breast cancer set ($n = 40$), with 82% sensitivity, 72% specificity and 75% predictive accuracy. Disease relapse was associated with significantly

Abbreviations: AUC, area under the curve; CERM, Center of Magnetic Resonance; CPMG, Carr–Purcell–Meiboom–Gill; ER, oestrogen receptor; GC, gas chromatography; ¹H NMR, proton nuclear magnetic resonance; HER2, human epidermal growth factor receptor 2; MS, mass spectroscopy; MSKCC, Memorial Sloan-Kettering Cancer Center; NOESY1D, nuclear overhauser effect spectroscopy pulse sequence; RF, random forest; ROC, receiver operating curves.

* Corresponding author. Tel.: +39 0574 802520.

E-mail addresses: tenori@cerm.unifi.it (L. Tenori), catherine.oakman@mh.org.au (C. Oakman), morrisp1@mskcc.org (P.G. Morris), gralka@cerm.unifi.it (E. Gralka), nhturner@usl4.toscana.it (N. Turner), scappadona@usl4.toscana.it (S. Cappadona), fornierm@mskcc.org (M. Fornier), hudisc@mskcc.org (C. Hudis), nortonl@mskcc.org (L. Norton), luchinat@cerm.unifi.it (C. Luchinat), adileo@usl4.toscana.it (A. Di Leo).

¹ Leonardo Tenori and Catherine Oakman have equally contributed to this work.

<http://dx.doi.org/10.1016/j.molonc.2014.07.012>

1574-7891/© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

lower levels of histidine ($p = 0.0003$) and higher levels of glucose ($p = 0.01$), and lipids ($p = 0.0003$), compared with patients with no relapse.

Conclusions: The performance of a serum metabolomic prognostic model for disease relapse in individuals with ER-negative early stage breast cancer is promising. A confirmation study is ongoing to better define the potential of metabolomics as a host and tumour-derived prognostic tool.

© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Breast cancer is the most common malignancy and most common cause of cancer death in women (Ferlay et al., 2010). There is marked heterogeneity in breast cancer biology and disease behaviour. Among individuals with seemingly similar disease as assessed by clinico-pathological features, immunohistochemistry and molecular platforms, outcomes can be substantially different.

The window of opportunity for curative intervention in breast cancer is in early stage disease. Following surgical excision of the breast lesion and surgical sampling and/or dissection of axillary nodes, patients might be offered loco-regional radiotherapy and/or post-operative (adjuvant) systemic therapy. The rationale behind this approach is that residual micrometastatic disease might be eradicated by chemotherapy, targeted anti-human epidermal growth factor receptor 2 (HER2) therapy and/or targeted endocrine therapy. If not eradicated, micrometastases might progress to incurable disseminated breast cancer.

In current clinical practice adjuvant therapy is indicated on the assumption of residual micrometastases. The primary disease is assessed using traditional clinico-pathologic features with or without gene profiling, and an estimation of the risk of recurrence is thus made (Paik et al., 2004; Ravdin et al., 2001). Micrometastatic disease is detectable, as circulating tumour cells in the peripheral blood and disseminated tumour cells in the bone marrow, although not all patients with micrometastases will develop clinically evident macrometastatic disease (Braun et al., 2005; Stathopoulou et al., 2002; Xenidis et al., 2003). Factors beyond the presence of micrometastases, such as tumour cell dormancy, host immunity, and the microenvironment, influence the clinical outcome.

Novel prognostic and predictive biomarkers may refine risk assessment and guide use of systemic therapy in individuals with early breast cancer. In this setting there are promising tools such as the various –omics, including metabolomics, a science dedicated to the global study of small molecules and metabolites (Nicholson, 2006). Metabolomics combines high resolution data-rich analytical methodology with advanced chemometric data interpretation. The ‘metabolome’, the extensive analysis of hundreds of metabolites in a biological specimen, can be considered the downstream end product of the complex interaction of genome, transcriptome and proteome. By its very nature of being downstream it may be a very sensitive tool for phenotype assessment.

The metabolome is affected by physiological, pathological and iatrogenic factors (Griffin, 2003; Urbanczyk-Wochniak

et al., 2007). Breast cancer has been associated with marked metabolic shifts, which have been demonstrated in many pre-clinical and clinical metabolomic studies of breast cells, breast tissue, serum and urine (Aboagye and Bhujwala, 1999; Budczies et al., 2012; Katz-Brull et al., 2002; Li et al., 2011; Mackinnon et al., 1997; Mountford et al., 2001; Singer et al., 1995). Metabolomics has been explored as a tool for diagnosis of breast cancer, refined sub-classification of breast cancer, and prediction of treatment sensitivity (Li et al., 2011; Mountford et al., 2001; Asiago et al., 2010; Borgan et al., 2010; Giskeodegard et al., 2012; Oakman et al., 2011; Slupsky et al., 2010).

A potential strength of serum and urine metabolomic analyses is that this approach provides a composite metabolomic snapshot of both the tumour and the host. By comparing samples from patients with early versus metastatic disease, host features conducive to tumour progression might be identified, with incorporation of both tumour and host factors from the outset.

In view of the need for more refined prognostic estimation in early breast cancer, we undertook this study to explore whether metabolomics can add prognostic information in individuals with early breast cancer. We assessed serum metabolomic profiles in breast cancer patients using proton nuclear magnetic resonance (^1H NMR) spectroscopy, with two hypotheses: (1) serum metabolomic profiles would be different between women following surgery for early breast cancer and women with metastatic disease, due to tumour-specific changes in the ^1H NMR detectable metabolomic profile; and (2) some patients with early breast cancer would be recognized by metabolomic analysis as having metastatic disease due to the presence of residual micrometastases.

We report serum metabolomic distinction between women following surgery for early breast cancer and women with metastatic disease, and further, we report metabolomic classification of a minority of early patients as metastatic, in whom future disease relapse was more likely.

2. Patients and methods

This was a collaborative project between the Breast Cancer Medicine Service, Memorial Sloan-Kettering Cancer Center (MSKCC), New York, United States; the Center of Magnetic Resonance (CERM), University of Florence, Sesto Fiorentino, Italy; and the ‘Sandro Pitigliani’ Medical Oncology Department, Hospital of Prato, Prato, Italy.

The study protocol was approved by the Institutional Ethics Committees at both MSKCC and the Hospital of Prato.

2.1. Patients

Serum samples were selected from the MSKCC breast cancer biobank. Patients attending the breast cancer medicine service at MSKCC who provide written informed consent and undergo standard serum biochemical analyses routinely have left over serum stored at -80°C for future research purposes. The Breast Cancer Medicine Service maintains a database of all clinico-pathological data and clinical outcomes.

This database was reviewed to identify cohorts of women with early and metastatic breast cancer. Eligible early breast cancer patients were required to have a post-operative serum sample, collected before starting adjuvant therapy and within 90 days of surgery, with either documented disease relapse or follow-up of at least five years with no relapse. Patients with relapse were included regardless of oestrogen receptor (ER) status. Conversely, the group without relapse was restricted to patients with ER-negative disease, as five years of follow-up was felt to be inadequate to exclude a relapse in ER-positive disease in which relapse may occur out to 10 years. As this was a retrospective study, all identified and suitable early breast cancer patients, based on the above criteria, were included. Eligible metastatic breast cancer patients were required to have a post-diagnosis serum sample, and were unrestricted with regards to ER status and duration of metastatic disease.

Study serum samples (500 μl) were maintained at -80°C from collection until transfer over dry ice from MSKCC to CERM, where they were again stored at -80°C until analysis. Serum samples and corresponding clinico-pathological data were anonymized prior to transfer.

2.2. ^1H NMR spectral acquisition

Frozen serum samples were thawed at room temperature and shaken before use. A phosphate sodium buffer (300 μl , 70 mM, pH 7.4) was added in a 1:1 ratio before analysis. The mixture was homogenized by vortexing for 30 s. 450 μl were transferred into a 4.25 mm NMR tube (Bruker BioSpin srl).

^1H NMR spectra for all samples were acquired using a Bruker 600 MHz metabolic profiler (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm CPTCI ^1H - $^{13}\text{C}/^{31}\text{P}$ - ^2H cryoprobe including a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic sample changer. A BTO 2000 thermocouple served for temperature stabilization at the level of approximately 0.1 K at the sample. Before measurement, samples were kept for at least three minutes inside the NMR probehead for temperature equilibration (300 K).

For each sample, three ^1H NMR spectra were acquired with water peak suppression: (i) a standard nuclear Overhauser effect spectroscopy pulse sequence (NOESY1Dpresat; Bruker), (ii) a Carr–Purcell–Meiboom–Gill (CPMG; Bruker) spin-echo sequence to suppress signals arising from high molecular weight molecules, (iii) a diffusion edited sequence (ledbpgppr2s1d; Bruker) with a diffusion time of 120 ms.

2.3. ^1H NMR spectral processing

Free induction decays were multiplied by an exponential function equivalent to a 1.0 Hz line-broadening factor before applying Fourier transform. Transformed spectra were manually corrected for phase and baseline distortions and calibrated (TMSP peak at 0.00 ppm) using TopSpin (Version 2.1, Bruker). Each 1D spectrum in the range between 0.02 and 10.00 ppm was segmented into 0.02 ppm chemical shift bins and the corresponding spectral areas were integrated using AMIX software (version 3.8.8, Bruker BioSpin). Regions between 4.5 and 6.0 ppm containing residual water signal were removed. The total spectral area was calculated on the remaining bins and data normalization was carried out prior to pattern recognition. Binning is a means to reduce the number of total variables and to compensate for small shifts in the signals, making the analysis more robust and reproducible (Holmes et al., 1994; Spraul et al., 1994). Using 0.02 ppm binning, the dimension of the system was reduced to 416 bins.

2.4. Statistical analysis

A statistical analysis plan was prepared before starting the analyses, based on the results of a prior study from our group (Oakman et al., 2011), where we first observed clusterization of serum metabolomic spectra from early and advanced breast cancer patients.

The size of the early breast cancer group was not predefined and was based on the availability of suitable cases with adequate clinical follow-up data and matched serum samples. In contrast, it was determined that a sample size of 95 patients with metastatic breast cancer would be expected to provide sufficient power (power = 0.9, $p < 0.05$) to detect a meaningful difference between this metastatic disease group and the early breast cancer group (Cohen's $d = 0.5$).

The group of early breast cancer patients was randomly split into two independent cohorts of equal patients number. Initial analyses were restricted to patients in the first cohort (hereafter referred to as the training set). The first step was to establish if serum metabolomic profiles could distinguish between patients with metastatic breast cancer and patients post-recent surgery for operable primary breast cancer. This analysis was unsupervised as to whether early breast cancer patients developed relapse.

For discriminative models, a Random Forest (RF) classifier (Breiman, 2001) was built to separate early breast cancer patients from patients with metastatic disease. RF is a classification algorithm that uses an ensemble of unpruned decision trees, each of which is built on a bootstrap sample of the training data using a randomly selected subset of variables. RF can deal with large numbers of predictor variables simultaneously, even in the presence of complex non linear interactions (Strobl et al., 2009) and it is almost immune from the overfitting due to the total number of variables in the data. This algorithm has many strengths in metabolomics: i) it is applicable when there are more variables than samples; ii) it is relatively insensitive to noise; iii) it allows visualization of data in a reduced discriminant space using the proximity matrix calculated during the process of forest growing; iv) the percentage of trees in the forest that assign one sample to a specific class can be interpreted

as a probability of class belonging; and v) it gives an unbiased estimate of the classification error using the out-of-bag samples, avoiding the need for time-consuming cross validation. The algorithm begins with the selection of many bootstrap samples from the original data. In a typical bootstrap sample, approximately 63% of the original observations occur at least once. Observations in the original data set that do not occur in a bootstrap sample are defined as out-of-bag (OOB) observations. A classification tree is fitted to each bootstrap sample but at each node only a small number of randomly selected variables (i.e. bins) are available for binary partitioning. The trees are fully grown, and each is used to predict the OOB observations. The predicted class of an observation is calculated by the majority vote of the OOB predictions for that observation. This classification rule was used to discriminate between early and metastatic patients. The percentage of the total number of trees in the forest that classified a sample as early or metastatic can be interpreted as the probability of that sample belonging to that group. The percentage of trees in the model classifying an early patient as metastatic was interpreted as a measure of the metabolic risk of that patient. Hence, for early patients, a score that expressed the probability of being metastatic was created and designated as the 'RF risk score'. For each patient, three 'RF risk scores' were derived using the three types of spectra. We assumed that higher scores correlated with higher risk of developing a relapse. Importantly, the calculated RF score is not a simple summation of the levels of several metabolites, but rather, is a complex multivariate entity created using as input the whole binned spectral data.

For all calculations, the R package 'Random Forest' (Liaw and Wiener, 2002) was used to grow a forest of 1000 trees, using the default settings. Separate models were built for NOESY1D, CPMG and Diffusion edited spectra. Statistical analyses were performed using the R statistical environment (Ihaka and Gentleman, 1996).

The second step was to test the hypothesis that metabolomic classification of some early patients as metastatic was due to metabolomic detection of signals from micrometastatic disease with likelihood for progression to metastatic disease. Using receiver operating characteristics (ROC) analysis, the performances of the three RF risk scores were compared with actual breast cancer outcome. Disease relapse was defined as loco-regional and/or distant breast cancer recurrence documented according to conventional clinico-radiological criteria, with pathology confirmation required in the case of equivocal findings.

A prognostic model was created using the CPMG RF risk score. The model was restricted to the CPMG score as this had the highest predictive value in the training set. CPMG RF risk scores were scaled to yield an integer score range of 0–100. Using actual follow-up data for patients who relapsed, a CPMG RF score cut-off was calculated in the training set by a judicious choice between sensitivity and specificity, then maintained as a constant when used for the prediction of relapse risk of all the samples in the 'validation' set. In addition, the CPMG RF risk score was evaluated in univariate and multivariate analyses firstly with tumour size, nodal status and age, and secondly with the Adjuvant!Online (<http://www.adjuvantonline.com>) relapse risk score, to assess the prognostic value of the metabolomic risk score in the context of known clinico-pathological variables.

2.5. Data reproduction in an independent set of patients

The process of choosing an appropriate threshold of the CPMG RF score was a supervised analyses, in the sense that creation and optimization of the model required information about actual relapse. For this reason, the results required testing in an independent set of early stage patients, designated for simplicity as the validation set. However, in this context the word 'validation' does not mean that results confirmed through the validation set should be considered definitive.

2.6. Metabolite analysis

The NMR spectrum from each sample was aligned with reference to trimethylsilylpropionic-acid at 0 ppm. Spectral regions within the range 0.5–9.0 ppm were analysed after excluding the region between 4.5 and 6.0 ppm that contained the water peak.

Differential metabolites levels between the metastatic and early cohorts were analysed using the non parametric Wilcoxon test. The intensities of the peaks were compared in the raw spectra (without normalization) and after adjustment for multiple testing using the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995) to control the false discovery rate. *p* values of <0.05 were considered significant.

3. Results

3.1. Patients

From the MSKCC serum bank, 95 women with metastatic breast cancer and 80 women with early operable breast cancer were identified. Stored serum samples had been collected between 2003 and 2010. Clinical and pathological characteristics, which are listed in Table 1, were obtained from the biobank database, with reference back to the patient's electronic medical record if clarification was required.

Most women with early breast cancer had received post-operative systemic therapy: 54 (68%) women received chemotherapy. Of note, only 9 of 24 HER-2 positive patients received trastuzumab-based adjuvant therapies. Three women (4%) had ER-positive disease and all three received adjuvant endocrine therapy without chemotherapy. A minority of early stage patients (*n* = 7 (9%)) had received neo-adjuvant chemotherapy. Inclusion of these patients was considered reasonable as all seven patients had a wash-out period of at least three weeks between last dose of chemotherapy and post-surgery blood sampling, and patients may still have had residual micrometastatic disease not eradicated by neo-adjuvant treatment.

Early breast cancer patients were randomly split into a training set (*n* = 40) and a validation set (*n* = 40). Of the 80 patients with early stage disease, 21 had documented disease relapse, 10 of whom were in the training set and 11 in the validation set.

Patients with metastatic disease were similar in age to those in the early breast cancer cohort. The majority of metastatic patients (62%) had ER-positive disease. Twenty-four % of patients had HER2 negative disease. The median time from diagnosis of metastatic disease until blood sample was 59 days (range 2–1737). Samples from metastatic breast cancer

Table 1 – Patient and tumour characteristics.

Characteristic	Metastatic breast cancer n = 95	Early breast cancer		
		All n = 80	Training set n = 40	Validation set n = 40
Age, mean (range) years	53 (28–80)	53 (31–88)	56 (31–88)	51 (32–78)
Tumour size				
<2 cm	NA	59 (74%)	29 (73%)	30 (75%)
≥2 cm		19 (24%)	10 (25%)	9 (23%)
Unknown		2 (2%)	1 (2%)	1 (2%)
Grade				
I	1 (1%)	0	0	0
II	11 (11%)	2 (3%)	1 (2%)	1 (2%)
III	69 (73%)	69 (86%)	34 (85%)	35 (88%)
Unknown	14 (15%)	9 (11%)	5 (13%)	4 (10%)
Lymph node involvement				
Node negative	NA	41 (51%)	19 (48%)	22 (55%)
Node positive		39 (49%)	21 (52%)	18 (45%)
ER				
Positive	59 (62%)	3 (4%)	2 (5%)	1 (2%)
Negative	33 (35%)	77 (96%)	38 (95%)	39 (98%)
Unknown	3 (3%)	0	0	0
HER2				
Positive	23 (24%)	24 (30%)	14 (35%)	10 (25%)
Negative	71 (75%)	55 (69%)	25 (63%)	30 (75%)
Unknown	1 (1%)	1 (1%)	1 (2%)	0
Neo-adjuvant chemotherapy	NA	7 (9%)	3 (8%)	4 (10%)
Adjuvant chemotherapy				
None	NA	16 (20%)	7 (17%)	9 (22%)
Chemotherapy		54 (68%)	27 (68%)	27 (68%)
Chemotherapy + anti-HER2 therapy		9 (11%)	5 (13%)	4 (10%)
Unknown		1 (1%)	1 (2%)	0
Endocrine therapy		3 (4%)	2 (5%)	1 (2%)

ER: oestrogen receptor; HER2: human epidermal growth factor receptor 2; NA: not applicable.

patients were drawn before starting a new line of therapy for advanced disease. Of the 95 patients with advanced disease, 56 were deceased at the time of the present analysis. Median overall survival for the deceased patients was 19 months (range 1–58 months).

3.2. Spectra

NMR spectra of samples were obtained using NOESY1D, CPMG and Diffusion editing, and showed clear signals for multiple metabolites and small molecules.

3.3. Training set

Supervised analysis using RF showed spectra clusterization from early and metastatic patients (Figure 1). Accuracy in predicting early or metastatic status was 83.7% (95% CI 83.6–83.9%) for CPMG, 86.7% (95% CI 86.4–86.8%) for NOESY1D, and 84.4% (95% CI 83.6–85.1%) for Diffusion editing.

3.4. Comparison between RF risk scores and actual relapse

Next, the hypothesis that metabolomic classification of some early patients as metastatic was due to the presence of residual micrometastatic disease was tested by comparing metabolomic RF risk scores and actual relapse (Figure 2). Using ROC

analyses, the best prediction for RF score was seen with the CPMG spectra, with area under the curve (AUC) of 0.863, compared with AUC 0.817 and 0.607 for NOESY1D and Diffusion editing spectra, respectively.

3.5. Relapse prediction by CPMG RF risk score

The RF score is a continuous variable. In order to use this approach to create a predictive model, a CPMG RF risk score threshold for relapse was required. Accuracy of the RF risk score was maximized using a threshold of ≥ 53 , which yielded sensitivity of 90% (95% CI 74.9–94.8%), specificity of 67% (95% CI 63.0–73.4%), and overall accuracy for predicting likelihood of relapse of 73% (95% CI 70.6–74.8%). For raw data, see Additional Table 1.

3.6. Data reproduction in an independent set of patients

The validation set was evaluated using an unsupervised analysis. Spectra of the validation samples were classified as either ‘metastatic’ or ‘early’ using the CPMG RF risk score model derived from the training set. Comparison between metabolomic classification and actual outcome demonstrated high correlation with AUC 0.824 (Figure 3). Using the CPMG RF risk score threshold ≥ 53 , sensitivity, specificity, and predictive accuracy were 82%, 72% and 75%, respectively.

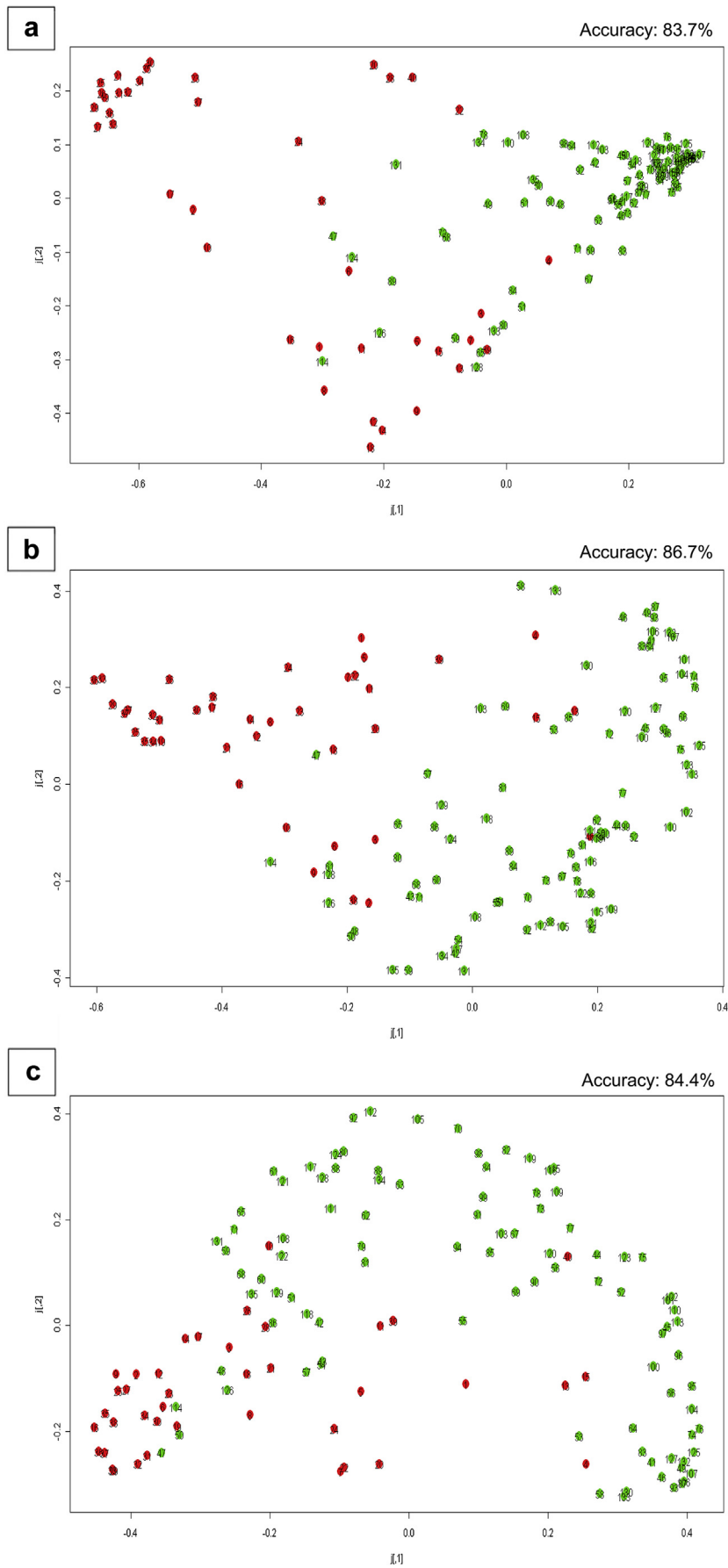


Figure 1 – Clusterization of serum metabolomic profiles. Discrimination between metastatic (green, $n = 95$) and early (red, $n = 40$) breast cancer patients using the random forest classifier. (a) CPMG; (b) NOESY1D; (c) Diffusion.

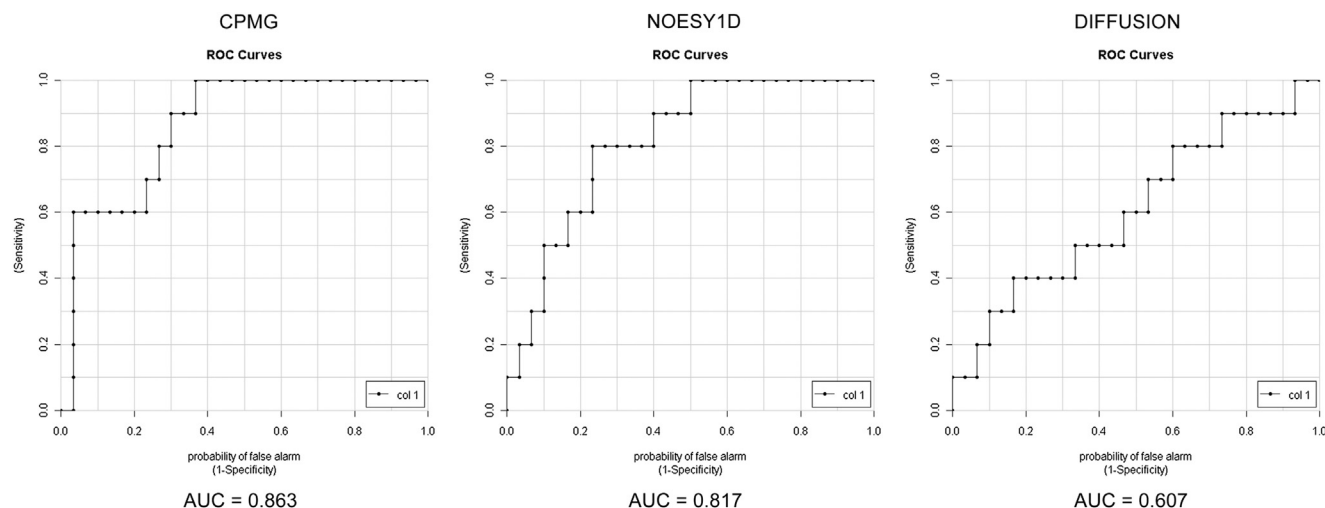


Figure 2 – Training set. Comparison between metabolomic classification and actual relapse. The receiver operator curves (ROC) and the area under the curve (AUC) scores are presented for CPMG, NOESY1D and Diffusion.

3.7. Comparison of RF risk score with known prognostic factors

The known prognostic factors age, tumour size and nodal status were compared with CPMG RF risk score in univariate and multivariate analyses. As all early stage patients had grade 3 tumours, histological grade was not included in the model. In a univariate analysis tumour size, nodal status and RF risk score were all significantly associated with relapse prediction. Coefficients on univariate regression were 1.08, $p = 0.008$ for tumour size; 0.69, $p = 0.011$ for nodal status; and 7.24, $p = 0.000138$ for RF risk score, while the coefficient for age was not statistically significant (-0.008 , $p = 0.68$). In the multivariate regression model, which included tumour size, nodal

status and RF score, tumour size and RF score showed a trend toward significance (coefficients of 0.70, $p = 0.062$, and 4.34, $p = 0.060$ for tumour size and RF risk score, respectively), while nodal status was no longer significantly associated with relapse (coefficient 0.40, $p = 0.32$), suggesting a moderate dependence between these variables.

As an alternative means of evaluating the independent prognostic value, the RF risk score was compared in univariate and multivariate models with relapse risk as predicted by Adjuvant!Online (AoL). AoL is an accepted and validated tool that provides prognostic data based upon a patient's clinicopathologic profile, incorporating data on age, comorbidities, tumour size, grade, nodal status and ER status. It provides estimates of 10-year risk of disease relapse and death. For the comparison with RF risk score, AoL risk score was calculated as the 10-year risk of cancer relapse (in absence of adjuvant therapy) using AoL (Standard Version 8.0). AoL risk score was significantly associated with relapse in the univariate model, with regression coefficient of 0.04, $p = 0.0038$. However, in a multivariate analysis that incorporated AoL risk score and RF risk score, only RF risk remained statistically significantly associated with relapse (coefficient: 4.64, $p = 0.03$), demonstrating that RF risk score has independent prognostic value.

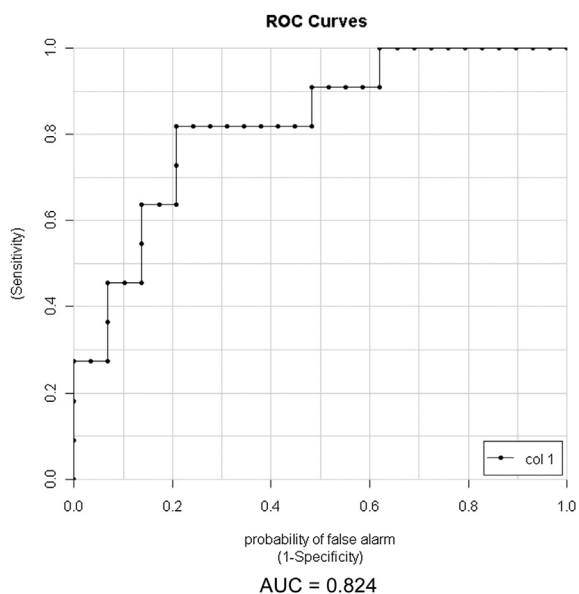


Figure 3 – Validation set. Comparison between CPMG random forest risk score metabolomic classification and actual relapse. The receiver operator curve (ROC) and the area under the curve (AUC) score are presented for the CPMG analysis.

3.8. Analysis of confounding factors

3.8.1. ER status

Due to the inclusion criteria for early stage patients, there was a difference in ER status between the early and metastatic patient groups. In order to test for any confounding effects of the ER status on the NMR profile, ER status prediction was attempted using the Random Forest approach. This approach was based on the principle that if no discrimination between ER-positive and ER-negative patients on evaluation of the metabolomic spectra was seen, it implied that no observable effects related to ER status are embedded in the serum NMR profile, thus excluding ER status as a confounder in the present analyses of the spectra. For prediction, samples from 92

metastatic patients were utilized (59 ER-positive, 33 ER-negative). Samples from three patients with unknown ER status were excluded. The overall accuracy for prediction of ER status was 51.2% (95% CI 50.9–51.5%, $p = 0.68$), with a sensitivity of 46.2% (45.6–46.7%), and a specificity of 54.0% (95% CI 53.7–54.2%). These data suggest that ER status was not encoded in the serum metabolome of patients and, therefore, that ER status should not be considered a confounding factor in the risk prediction analyses.

Furthermore, the analyses comparing spectra of early and metastatic patients was repeated including only the ER-negative patients from both groups ($n = 77$, and $n = 33$ in the early and metastatic groups, respectively), and showed comparable results to the original analysis that incorporated all patients irrespective of ER status. Overall accuracy was 82.4% (95% CI 82.3–82.6%), with a sensitivity of 82.4% (95% CI 82.2–82.6%), and a specificity of 82.5% (95% CI 82.3–82.7%), demonstrating again that a significant metabolic difference does exist between the metastatic and early cohorts.

3.8.2. Timing of serum sample

The time interval between surgery and the date of blood sampling, which ranged from five to 80 days in the early breast cancer cohort, was also explored as a potential confounder. A model using the Random Forest classification was built to evaluate if time interval between surgery and serum sample collection could be predicted from the metabolomic data. For this purpose a dichotomous variable was created, dividing early samples into two categories: an interval of less than 30 days between surgery and collection and an interval of at least 30 days. No significant discrimination in metabolomic spectra was seen based on this variable (accuracy = 53.6%, 95% CI 53.5–53.8%), supporting lack of confounding by time interval between surgery and serum sample collection.

3.9 Metabolites

An analysis of the NMR spectra was conducted to identify which metabolites were contributing to the metastatic profiles. These metabolites were compared with metabolites contributing to the profiles from early patients. All metastatic patients and all early patients were included in these analyses. Relative serum concentrations of metabolites were estimated through integration of the signals in the NMR spectra and the comparison was performed using univariate Wilcoxon test. Compared with profiles from early patients, serum profiles from patients with metastatic disease had significantly lower levels of histidine and significantly higher serum levels of glucose, lactate, tyrosine and lipids. After correcting for multiple testing, differences in tyrosine and lactate levels were no longer statistically significant, while the other three metabolites remained significantly different. [Figure 4](#) depicts the discriminant metabolites and the associated unadjusted and adjusted p values.

a metabolomic signal associated with disease recurrence in individuals with early breast cancer.

We observed distinction between serum metabolomic profiles of early and metastatic breast cancer patients. This observation is consistent with results of our previous study, where we found a discrimination between patients with early disease, whose blood sample was drawn pre-operatively, and patients with metastatic disease, with sensitivity of 75%, specificity of 69%, and predictive accuracy of 72% ([Oakman et al., 2011](#)), and with recently published data from Jobard et al., who similarly demonstrated significant differentiation between serum metabolomic profiles of early and metastatic breast cancer patients ([Jobard et al., 2014](#)).

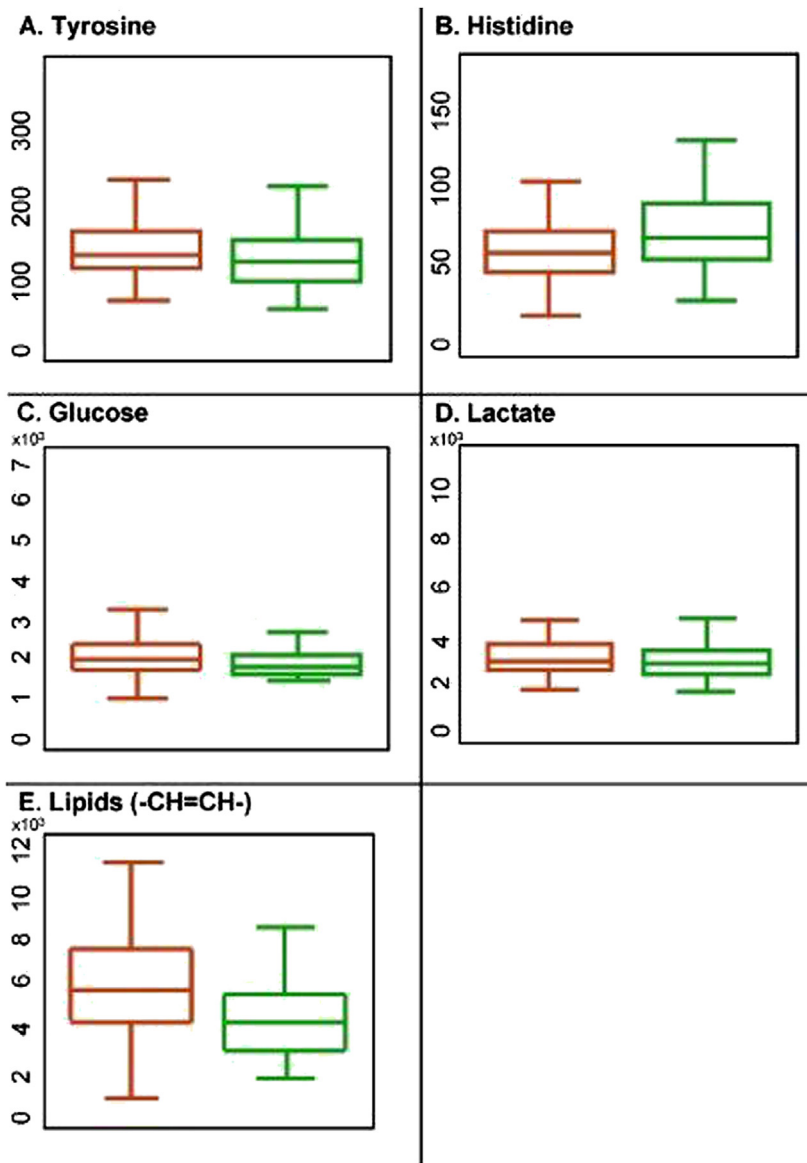
Furthermore in this current study we observed the promising performance by metabolomics for identification of early patients with and without subsequent relapse. In particular, we identified a CPMG RF signature in a training set of early breast cancer patients, with the predictive utility of this model reproduced in an independent set of patients. Clinical studies clearly demonstrate that, while adjuvant chemotherapy improves disease free survival, many patients treated with surgery alone remain disease free in the long term. After 30 years of follow-up in node-positive disease and 20 years of follow-up in node-negative ER-negative disease, approximately one quarter and one half of patients, respectively, were disease free after surgery alone ([Bonadonna et al., 2005](#)). There is still limited clinical capacity to identify these individuals who do not require, and obtain no benefit from, adjuvant intervention. With further validation, incorporation of this approach into relapse risk assessment might allow identification of patients with low metabolomic risk of relapse, who have been cured by surgery alone, and thus who might be spared adjuvant treatment and its associated toxicities.

Patients with subsequent disease relapse have the most to potentially gain from adjuvant systemic therapy. A strength of our study is that serum samples were collected prior to adjuvant therapy and within a short period post-operatively (5–80 days), a clinically meaningful time for making adjuvant treatment decisions. A recent study from Asiago et al. used a combination of NMR and gas chromatography–mass spectroscopy (GC–MS) to identify a serum metabolomic signal for early detection of metastatic disease in individuals undergoing surveillance following early breast cancer ([Asiago et al., 2010](#)). These results are intriguing; however metastatic breast cancer is incurable, whether diagnosed early or late and there currently is no advantage to early detection of low volume, asymptomatic metastatic disease. In contrast, detection of a post-operative signal of micrometastatic disease and administration of curative-intent adjuvant systemic treatment might alter clinical outcomes.

Many metabolites have been shown to correlate with breast cancer development and progression ([Aboagye and Bhujwala, 1999](#); [Singer et al., 1995](#)). Marked changes are reported in cellular phospholipid metabolism, glycolysis and amino acid metabolism ([Aboagye and Bhujwala, 1999](#); [Katz-Brull et al., 2002](#); [Li et al., 2011](#); [Griffin and Shockcor, 2004](#)). ^1H NMR metabolomic profiles contain qualitative and quantitative information on hundreds of metabolites and small molecules ([Aranibar et al., 2011](#); [Serkova and Niemann, 2006](#)). A strength of non-targeted global spectrum analysis over

4. Discussion

In this study we employed serum ^1H NMR spectral profiling and advanced chemometric data analysis methods to identify



Metabolite	Metastatic patients	Early patients	Raw P-values	Adjusted P-values
Tyrosine	128	120	0.0480	0.2496
Histidine	71	81	2.02e-5	0.0003
Glucose	2023	1825	0.0012	0.0104
Lactate	19684	17070	0.0130	0.0845
Lipids (-CH=CH-)	5710	4167	2.13e-5	0.0003

Figure 4 – Discriminant metabolites. Discriminant metabolites ($p < 0.05$) between profiles from early (green, $n = 80$) and metastatic (red, $n = 95$) breast cancer patients. Box and whisker plots: horizontal line within the box = mean; bottom and top lines of the box = 25th and 75th percentiles, respectively; bottom and top whiskers = 5th and 95th percentiles, respectively. Median values (arbitrary units) are provided in the associated table, along with raw p values and p values adjusted for multiple testing. pts: patients.

targeted analysis of specific metabolites is the lack of need to make any assumptions on the identity of the metabolites that are relevant for the selected pathology. Non-targeted profiles capture a downstream readout of genetic signal, post-

genomic signalling, cross-talk between signalling pathways, and environmental influence. The complex metabolomic signal that correlates with disease relapse is presumed to contain metabolites relevant to residual micrometastatic

disease and metabolites relevant to the host's systemic response to the tumour.

In the current study the RF risk model used the entire NMR spectra. Subsequently the spectra were interrogated to identify key discriminating metabolites. Patients with metastatic disease were characterized by lower serum levels of histidine and higher serum levels of glucose, lactate, tyrosine and lipids, compared with post-operative early breast cancer subjects who did not relapse. These results are consistent with recent findings of Asiago et al., who also observed a reduction in serum histidine and an increase in tyrosine and lactate in women with a recurrent breast cancer, with respect to women with a history of primary breast cancer without recurrence (Asiago et al., 2010). Similarly, Jobard et al. demonstrated higher levels of histidine in early compared with metastatic breast cancer patients, while levels of several metabolites including phenylalanine, pyruvate, glutamate and glycerol were relatively higher in metastatic patients. Lipid levels were also elevated, although this difference did not reach statistical significance (Jobard et al., 2014). High lipids have been associated with high tumour cell proliferation, high cell membrane turnover and lipid activity in intracellular signal transduction (Aboagye and Bhujwalla, 1999; Katz-Brull et al., 2002; Cuadrado et al., 1993). Increased serum NMR intensity for lipid signals has also been attributed to inflammatory response in cancer (Bertini et al., 2012). We found significantly higher intensity of lipid signal resonances in patients with metastatic disease which may, at least in part, reflect a non-specific inflammatory response. Further cross-study comparison of specific metabolites is limited due to diversity in biospecimen (tissue, urine and serum), metabolomic platform (NMR and/or MS) and chemometric approach in published metabolomic literature in breast cancer to date.

One of the great challenges of metabolomic analyses is variation: innate physiological variability for individuals and biological variability for tumours. For any one person the profile may change with diet, diurnal rhythm, body flora, drugs, pathology and exercise. Between individuals, there is metabolic variation based on age, gender, race and hormonal status (Bollard et al., 2005). In this study, serum samples were sourced from one hospital and a single sample was analysed for each patient. In future studies it will be important to collect more than one sample per patient and to standardize the collection procedures (for instance, by defining a specific time of the day for blood samples drawing, preferably before any food and/or medication intake). In addition, in future studies it will be important to have samples collected in a limited time frame, while in the presented study samples were collected between years 2003 and 2010.

There are limitations of our study. Firstly, patient numbers are limited. We included all suitable early breast cancer patients identified in the MSKCC serum biobank and clinical database, and moreover we were able to reproduce results obtained in the training set in a small but independent 'validation' cohort. Nonetheless, further evaluation of the RF risk model in a larger patient cohort is required before any incorporation of a metabolomic risk model into clinical practice. Furthermore, the majority of early breast cancer patients in this study (79%) received adjuvant systemic therapy. This is a potential confounding factor, as chemotherapy may have

changed the outcome predicted from the post-operative, pre-chemotherapy sample. Early stage patients who were classified as metastatic by the metabolomic prognostic model may have been cured by adjuvant chemotherapeutic eradication of micrometastatic disease, in which case the utility of the model would be underestimated. Conversely, patients classified as early stage by the model may have had residual disease that was subsequently eradicated, a scenario where the prognostic value of the model would be overestimated. This can be particularly relevant in the subset of HER-2 positive patients who received trastuzumab-based adjuvant therapies.

In the current study, the analyses focused on early patients with ER-negative disease, with ER-negative disease comprising 96% of the early stage cohort. Patients with ER-positive disease were excluded unless they had relapsed within five years, due to concern that follow-up of five years was not of adequate duration to evaluate relapses in ER-positive disease, where recurrences may occur out to ten years. Importantly, we evaluated the performance of the RF risk model in the early stage patients excluding the three patients with ER-positive disease and found no significant difference in results (data not shown). Furthermore, metabolomic spectra showed no significant clusterization based on ER status, suggesting that ER status, while being clinically relevant, did not unduly influence the utility of the metabolomic risk model, and that differences in ER status between the early and metastatic disease groups did not confound the results obtained. Nonetheless, it will be important to explore whether the results from this prognostic model are reproducible in patients with ER-positive early breast cancer. A follow-up study has now been commenced, which aims to evaluate the metabolomic risk model in a much larger cohort of women with early stage, and ER-positive, breast cancer, thus addressing two limitations of the current study.

It is also of interest to explore how metabolomic approaches might be integrated with other prognostic information derived from the host, tumour and micrometastatic disease. Based on this concept, in a separate ongoing study, we are exploring whether there is prognostic synergy between metabolomic serum analyses and genomic assessment of the tumour.

5. Conclusions

The complexity of breast cancer biology and behaviour suggests that a multi-platform approach to risk assessment in early breast cancer is preferable to optimize risk prediction accuracy. In this study we have identified a serum metabolomic prognostic model for prediction of disease relapse in individuals with early stage breast cancer, that appears to have independent prognostic value over known prognostic clinicopathological variables. Importantly, results of this study should be seen as exploratory, due to the aforementioned limitations such as retrospective design, limited sample size, and lack of data relating to ER-positive early stage patients. Nonetheless, our results are promising as they suggest, for the first time, that detection of host and/or tumour-derived metabolomic signals assessed on serum samples taken before adjuvant

therapy may be informative with regard to individual patient outcomes. A validation study is ongoing, which aims to better define the potential of metabolomics as a host and tumour-derived prognostic tool.

Conflict of interest

The authors declare that they have no competing interests.

Authors' contributions

LT, CO, PGM, SC, MF, CL and AD participated in the design and coordination of the study. LT, CO and SC determined patient eligibility. PGM and MF identified eligible patients. LT and EG carried out the serum analysis and metabolomic spectra acquisition. LT, CO, SC, CH, NT and AD participated in the random forest analysis and interpretation. LT, CO, PGM, MF, NT and AD drafted the manuscript. CH, NT and LN provided critical expert revision of draft manuscript. All authors revised the manuscript critically and approved the final version for submission.

Acknowledgements

The authors wish to acknowledge the research support from the Breast Cancer Research Foundation (BCRF), New York, USA, the 'Sandro Pitigliani' Foundation, Prato, Italy, and the Associazione Italiana per la Ricerca sul Cancro (AIRC), Milan, Italy.

The funding institutions were not involved in the study design, in the collection/analysis/interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Appendix A. Supplementary material

Supplementary data related to this article can be found online at <http://dx.doi.org/10.1016/j.molonc.2014.07.012>.

REFERENCES

- Aboagye, E.O., Bhujwala, Z.M., 1999. Malignant transformation alters membrane choline phospholipid metabolism of human mammary epithelial cells. *Cancer Res.* 59, 80–84.
- Aranibar, N., Borys, M., Mackin, N.A., Ly, V., Abu-Absi, N., Abu-Absi, S., Niemitz, M., Schilling, B., Li, Z.J., Brock, B., Russell 2nd, R.J., Tymiak, A., Reilly, M.D., 2011. NMR-based metabolomics of mammalian cell and tissue cultures. *J. Biomol. NMR* 49, 195–206.
- Asiago, V.M., Alvarado, L.Z., Shanaiah, N., Gowda, G.A., Owusu-Sarfo, K., Ballas, R.A., Raftery, D., 2010. Early detection of recurrent breast cancer using metabolite profiling. *Cancer Res.* 70, 8309–8318.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Bertini, I., Cacciatore, S., Jensen, B.V., Schou, J.V., Johansen, J.S., Kruhoffer, M., Luchinat, C., Nielsen, D.L., Turano, P., 2012. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res.* 72, 356–364.
- Bollard, M.E., Stanley, E.G., Lindon, J.C., Nicholson, J.K., Holmes, E., 2005. NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* 18, 143–162.
- Bonadonna, G., Moliterni, A., Zambetti, M., Daidone, M.G., Pilotti, S., Gianni, L., Valagussa, P., 2005. 30 years' follow up of randomised studies of adjuvant CMF in operable breast cancer: cohort study. *BMJ* 330, 217.
- Borgan, E., Sitter, B., Lingjaerde, O.C., Johnsen, H., Lundgren, S., Bathen, T.F., Sorlie, T., Borresen-Dale, A.L., Gribbestad, I.S., 2010. Merging transcriptomics and metabolomics – advances in breast cancer profiling. *BMC Cancer* 10, 628.
- Braun, S., Vogl, F.D., Naume, B., Janni, W., Osborne, M.P., Coombes, R.C., Schlimok, G., Diel, I.J., Gerber, B., Gebauer, G., Pierga, J.Y., Marth, C., Oruzio, D., Wiedswang, G., Solomayer, E.F., Kundt, G., Strobl, B., Fehm, T., Wong, G.Y., Bliss, J., Vincent-Salomon, A., Pantel, K., 2005. A pooled analysis of bone marrow micrometastasis in breast cancer. *N. Engl. J. Med.* 353, 793–802.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Budczies, J., Denkert, C., Muller, B.M., Brockmoller, S.F., Klauschen, F., Gyorffy, B., Dietel, M., Richter-Ehrenstein, C., Marten, U., Salek, R.M., Griffin, J.L., Hilvo, M., Oresic, M., Wohlgemuth, G., Fiehn, O., 2012. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue – a GC-TOFMS based metabolomics study. *BMC Genomics* 13, 334.
- Cuadrado, A., Carnero, A., Dolfi, F., Jimenez, B., Lacal, J.C., 1993. Phosphorylcholine: a novel second messenger essential for mitogenic activity of growth factors. *Oncogene* 8, 2959–2968.
- Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127, 2893–2917.
- Giskeodegard, G.F., Lundgren, S., Sitter, B., Fjosne, H.E., Postma, G., Buydens, L.M., Gribbestad, I.S., Bathen, T.F., 2012. Lactate and glycine-potential MR biomarkers of prognosis in estrogen receptor-positive breast cancers. *NMR Biomed.* 25, 1271–1279.
- Griffin, J.L., 2003. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr. Opin. Chem. Biol.* 7, 648–654.
- Griffin, J.L., Shockcor, J.P., 2004. Metabolic profiles of cancer cells. *Nat. Rev. Cancer* 4, 551–561.
- Holmes, E., Foxall, P.J., Nicholson, J.K., Neild, G.H., Brown, S.M., Beddell, C.R., Sweatman, B.C., Rahr, E., Lindon, J.C., Spraul, M., Neidig, P., 1994. Automatic data reduction and pattern recognition methods for analysis of ¹H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal. Biochem.* 220, 284–296.
- Ihaka, R., Gentleman, R.R., 1996. A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314.
- Jobard, E., Pontoizeau, C., Blaise, B.J., Bachelot, T., Elena-Herrmann, B., Tredan, O., 2014. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. *Cancer Lett.* 343, 33–41.
- Katz-Brull, R., Seger, D., Rivenson-Segal, D., Rushkin, E., Degani, H., 2002. Metabolic markers of breast cancer: enhanced choline metabolism and reduced choline-ether-phospholipid synthesis. *Cancer Res.* 62, 1966–1970.

- Li, M., Song, Y., Cho, N., Chang, J.M., Koo, H.R., Yi, A., Kim, H., Park, S., Moon, W.K., 2011. An HR-MAS MR metabolomics study on breast tissues obtained with core needle biopsy. *PLoS ONE* 6, e25563.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R. News* 2, 18–22.
- Mackinnon, W.B., Barry, P.A., Malycha, P.L., Gillett, D.J., Russell, P., Lean, C.L., Doran, S.T., Barraclough, B.H., Bilous, M., Mountford, C.E., 1997. Fine-needle biopsy specimens of benign breast lesions distinguished from invasive cancer ex vivo with proton MR spectroscopy. *Radiology* 204, 661–666.
- Mountford, C.E., Somorjai, R.L., Malycha, P., Gluch, L., Lean, C., Russell, P., Barraclough, B., Gillett, D., Himmelreich, U., Dolenko, B., Nikulin, A.E., Smith, I.C., 2001. Diagnosis and prognosis of breast cancer by magnetic resonance spectroscopy of fine-needle aspirates analysed using a statistical classification strategy. *Br. J. Surg.* 88, 1234–1240.
- Nicholson, J.K., 2006. Global systems biology, personalized medicine and molecular epidemiology. *Mol. Syst. Biol.* 2, 52.
- Oakman, C., Tenori, L., Claudino, W.M., Cappadona, S., Nepi, S., Battaglia, A., Bernini, P., Zafarana, E., Saccenti, E., Fournier, M., Morris, P.G., Biganzoli, L., Luchinat, C., Bertini, I., Di Leo, A., 2011. Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. *Ann. Oncol.* 22, 1295–1301.
- Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., Hiller, W., Fisher, E.R., Wickerham, D.L., Bryant, J., Wolmark, N., 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* 351, 2817–2826.
- Ravdin, P.M., Siminoff, L.A., Davis, G.J., Mercer, M.B., Hewlett, J., Gerson, N., Parker, H.L., 2001. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J. Clin. Oncol.* 19, 980–991.
- Serkova, N.J., Niemann, C.U., 2006. Pattern recognition and biomarker validation using quantitative ¹H-NMR-based metabolomics. *Expert Rev. Mol. Diagn.* 6, 717–731.
- Singer, S., Souza, K., Thilly, W.G., 1995. Pyruvate utilization, phosphocholine and adenosine triphosphate (ATP) are markers of human breast tumor progression: a ³¹P- and ¹³C-nuclear magnetic resonance (NMR) spectroscopy study. *Cancer Res.* 55, 5140–5145.
- Slupsky, C.M., Steed, H., Wells, T.H., Dabbs, K., Schepansky, A., Capstick, V., Faught, W., Sawyer, M.B., 2010. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer Res.* 16, 5835–5841.
- Spraul, M., Neidig, P., Klauck, U., Kessler, P., Holmes, E., Nicholson, J.K., Sweatman, B.C., Salman, S.R., Farrant, R.D., Rahr, E., Beddel, C.R., Lindon, J.C., 1994. Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples. *J. Pharm. Biomed. Anal.* 12, 1215–1225.
- Stathopoulou, A., Vlachonikolis, I., Mavroudis, D., Perraki, M., Kouroussis, C., Apostolaki, S., Malamos, N., Kakolyris, S., Kotsakis, A., Xenidis, N., Reppa, D., Georgoulis, V., 2002. Molecular detection of cytokeratin-19-positive cells in the peripheral blood of patients with operable breast cancer: evaluation of their prognostic significance. *J. Clin. Oncol.* 20, 3404–3412.
- Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348.
- Urbanczyk-Wochniak, E., Willmitzer, L., Fernie, A.R., 2007. Integrating profiling data: using linear correlation to reveal coregulation of transcript and metabolites. *Methods Mol. Biol.* 358, 77–85.
- Xenidis, N., Vlachonikolis, I., Mavroudis, D., Perraki, M., Stathopoulou, A., Malamos, N., Kouroussis, C., Kakolyris, S., Apostolaki, S., Vardakis, N., Lianidou, E., Georgoulis, V., 2003. Peripheral blood circulating cytokeratin-19 mRNA-positive cells after the completion of adjuvant chemotherapy in patients with operable breast cancer. *Ann. Oncol.* 14, 849–855.