TRANSLATION ALPROTEOMICS 2 (2014) 52–59

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://www.elsevier.com/locate/trprot

Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer

Johanna Sonntag[a], Christian Bender[a,1], Zita Soons[b,c,2], Silvia von der Heyde[d], Rainer König[b,c,e], Stefan Wiemann[a], Hans-Peter Sinn[f], Andreas Schneeweiss[g], Tim Beißbarth[d], Ulrike Korf[a,*]

[a]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany
[b]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[c]Network Modeling, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute Jena, Jena, Germany
[d]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[e]Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany
[f]Department of Pathology, University of Heidelberg, Heidelberg, Germany
[g]National Center for Tumor Diseases (NCT), University of Heidelberg, Heidelberg, Germany

ARTICLE INFO

Article history:
Received 20 December 2013
Received in revised form 29 January 2014
Accepted 2 February 2014

Keywords:
Luminal breast cancer
Risk classification
Biomarker
Reverse phase protein array
Classification algorithm

ABSTRACT

A robust subclassification of luminal breast cancer, the most common molecular subtype of human breast cancer, is crucial for therapy decisions. While a part of patients is at higher risk of recurrence and requires chemo-endocrine treatment, the other part is at lower risk and also poorly responds to chemotherapeutic regimens. To approximate the risk of cancer recurrence, clinical guidelines recommend determining histologic grading and abundance of a cell proliferation marker in tumor specimens. However, this approach assigns an intermediate risk to a substantial number of patients and in addition suffers from a high interobserver variability. Therefore, the aim of our study was to identify a quantitative protein biomarker signature to facilitate risk classification. Reverse phase protein arrays (RPPA) were used to obtain quantitative expression data for 128 breast cancer relevant proteins in a set of hormone receptor-positive tumors ($n = 109$). Proteomic data for the subset of histologic G1 ($n = 14$) and G3 ($n = 22$) samples were used for biomarker discovery serving as surrogates of low and high recurrence risk, respectively. A novel biomarker selection workflow based on combining three different classification methods identified caveolin-1, NDKA, RPS6, and Ki-67 as top candidates. NDKA, RPS6, and Ki-67 were expressed at elevated levels in high risk tumors whereas caveolin-1 was observed as downregulated. The identified biomarker signature was subsequently analyzed using an independent test set (AUC = 0.78). Further evaluation of the identified biomarker panel by Western blot and mRNA profiling confirmed the proteomic signature obtained by RPPA. In conclusion, the biomarker signature introduced supports RPPA as a tool for cancer biomarker discovery.

1. Introduction

Breast cancer, the most frequent cancer entity among women, is nowadays recognized as a heterogeneous disease in terms of tumor morphology as well as at the molecular level [1–3]. Treatment of breast cancer patients with similar clinicopathologic features can result in different outcomes regarding disease progression and survival. Over the last few years, gene expression profiling has provided insights into molecular mechanisms associated with observed heterogeneous clinical outcome [4]. The seminal work of Sorlie and Perou identified intrinsic molecular subtypes, termed luminal A, luminal B, basal-like, and HER2-enriched, with unique biological and prognostic features [5]. The largest group of breast cancer patients suffers from luminal breast cancer with overexpression of hormone receptors as molecular hallmark. Luminal breast cancer comprises patients of the luminal A subtype with good prognosis whereas patients of the luminal B subtype are at a higher risk to suffer from recurrence [6]. Treatment of patients in these two groups is fundamentally different, with patients at higher recurrence risk requiring chemo-endocrine treatment, whereas others do not benefit from chemotherapy. Hence,

[1] Present address: TRON – Translational Oncology at the Johannes Gutenberg University of Mainz Medicine, Mainz, Germany.
[2] Present address: Department of Knowledge Engineering, Maastricht University, Maastricht, The Netherlands.
* *Corresponding author at*: German Cancer Research Center, Division of Molecular Genome Analysis B050, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.
E-mail address: u.korf@dkfz.de (U. Korf).

http://dx.doi.org/10.1016/j.trprot.2014.02.001

to avoid over- or under-treatment of patients with luminal breast cancer, tools allowing a clear-cut distinction of low and high risk are required. Although different approaches employing gene expression signatures or protein-based assays were introduced [7–11], a robust assessment of the recurrence risk in luminal breast cancer has remained a challenge.

To differentiate between low and high risk tumors, proliferation rate has emerged as a prominent feature, mainly supported by gene expression profiling data [4,12]. This is in line with information provided by histologic grade which is beside age, tumor size, and lymph node status a well-established independent prognostic factor, combining information on tumor proliferation and differentiation status. The Nottingham grading system is based on a semi-quantitative evaluation of morphologic tumor characteristics, in detail, tubule or gland formation, nuclear pleomorphism, and mitotic count, features which essentially reflect tumor proliferation and/or differentiation [13]. Based on the resulting score, tumor samples are assigned to three different categories, either well-differentiated (grade 1/G1), moderately differentiated (grade 2/G2) or poorly differentiated (grade 3/G3) [14]. For patients whose tumors were characterized as G1 or G3, prognostic information is univocal, with a good prognosis for G1 and a poor prognosis for G3 patients. However, a considerable percentage of patients are classified as G2 and in this instance a histologic grading provides no helpful information for treatment decisions.

In recent years, reverse phase protein arrays (RPPA) have emerged as a powerful high-throughput approach for targeted proteomics [15]. As a major advantage, RPPA allows to assess target protein expression quantitatively in large sample sets while requiring only a very low amount of biological sample [16] making this platform attractive for the analysis of clinical materials and biomarker discovery [17–19].

The objective of our study was to identify a robust protein signature using RPPA as a technical platform for targeted proteomics to assess the risk of cancer recurrence for breast cancer patients whose tumors had been diagnosed with histologic G2. Quantitative protein expression data were generated for 128 breast cancer relevant target proteins analyzing a set of 109 hormone receptor-positive tumors. A novel bioinformatics workflow combining three different classification algorithms was used to analyze RPPA data of histologic G1 and G3 tumors serving as surrogates of low and high risk breast cancer, respectively. The RPPA-derived signature was first subjected to an independent evaluation employing Western blot and mRNA profiling essentially confirming findings made by RPPA. Finally, the biomarker marker profile was translated into a risk classification score named R2LC suitable to predict the recurrence risk in single samples and validated using an independent test set comprising hormone-receptor positive tumors.

2. Materials and methods

2.1. Patient and sample characteristics

Tumor specimens (discovery set, $n = 109$) from patients diagnosed with primary invasive breast carcinoma were collected at the time of surgery between 2008 and 2010 at the Department of Gynecology and Obstetrics/National Center for Tumor Diseases, Heidelberg. None of the patients had received neoadjuvant therapy. Institutional Review Board approval was received as ethics vote no. S039/2008 and informed consent was obtained from all the patients. Tumor specimens were processed within 20 min after surgery. Samples were stored snap frozen at −80 °C until further use. Tumor specimens of the test set ($n = 145$) were obtained from the Tissue Bank of the National Center for Tumor Diseases (Heidelberg). Both sample sets comprised only tumors with >70% tumor cells and positive estrogen receptor status (immunoreactive score ≥3) as assessed by routine immunohistochemistry. Additional information on patient characteristics is summarized in Supplementary Tables S1, S2, and S3.

2.2. Reverse phase protein arrays

Frozen tumor samples were homogenized using a bead mill (TissueLyser, Qiagen) and tissue protein extraction reagent (T-PER, Thermo Scientific) supplemented with 1 mM EDTA, 5 mM NaF, 2 μM staurosporine, PhosSTOP Phosphatase Inhibitor Cocktail (Roche Applied Science), and Complete Mini Protease Inhibitor Cocktail (Roche Applied Science). Total protein concentration was determined by bicinchoninic acid assay (Thermo Scientific). Prior to spotting, tumor lysates were mixed with 4× SDS sample buffer (10% glycerol, 4% SDS, 10 mM DTT, 125 mM Tris–HCl, pH 6.8) and boiled for 5 min at 95 °C. Tumor lysates (total protein concentration 2 μg/μl) and dilution series of tumor sample pools serving as controls were spotted as technical triplicates and four identical subarrays on nitrocellulose-coated glass slides (Oncyte Avid, Grace-Biolabs) using a contact spotter (Aushon BioSystems). Slides were blocked with blocking buffer for fluorescent applications (Rockland Immunochemicals) in TBS (50%, v/v) containing 5 mM NaF and 1 mM Na_3VO_4 for 2 h at RT, prior to incubation with target-specific primary antibodies at 4 °C over night (Supplementary Table S4). Primary antibodies ($n = 128$) were selected to recognize proteins involved in major cancer signaling pathways with a special focus on breast cancer biology. Only highly target-specific antibodies were used and their validation was carried out as previously described [20]. Detection of primary antibodies was done with Alexa Fluor 680 F(ab′)2 fragments of goat anti-mouse IgG or anti-rabbit IgG in 1:8000 dilution (Life Technologies). In addition, representative slides were stained for total protein quantification using the protein dye Fast Green FCF as described before [21]. Images of all slides were obtained at an excitation wavelength of 685 nm and a resolution of 21 μm using the Odyssey Scanner (LI-COR). Signal intensities of each individual spot were quantified using GenePixPro 5.0 (Molecular Devices). Data preprocessing and quality control were performed with the R-package RPPanalyzer [22]. RPPA data of the discovery and the test cohort have been deposited in NCBI's Gene Expression Omnibus [23] and are accessible through GEO series accession number GSE47066 and GSE50861, respectively.

2.3. Biomarker selection process bootfs

We set up a biomarker (feature) selection workflow including three different algorithms for classification (SCAD-SVM: support vector machines using smoothly clipped absolute deviation penalty; RF-Boruta: random forests using the Boruta algorithm for feature selection; PAM: prediction analysis for microarrays utilizing the nearest shrunken centroid classifier [24–26]). We implemented the software in the R programming language and made it available through the *bootfs* R-package (https://r-forge.r-project.org/projects/bootfs/). The feature selection workflow was implemented as a bootstrapping procedure with 100 iterations as illustrated in Fig. 1 to derive a final feature set. Parameters for the SCAD-SVM method were set to 1000 maximum iterations and 500 minimum evaluations. The n.threshold parameter for the PAM classification was set to 30, and the maxRuns parameter for the RF-Boruta algorithm to 300. All other parameters were set to default values (for a detailed description of the parameter settings, refer to the documentation of the *bootfs* package). Abundance and co-occurrence of selected features were visualized graphically as network, termed the importance graph in the *bootfs* package. Parameters were set to vlabel.cex = 6, max_node_cex = 20, node.filter = 17, vlabel.cex.min = 0.8, vlabel.cex.max = 4, filter = 17, ewprop = 1.4, max_edge_cex=15.

2.4. Development of risk classification score R2LC

A decision rule was defined for the risk classification by setting up a logistic regression model for classifying the histologic grade depending on the protein expression levels of the selected biomarkers.