Identification of novel biomarkers associated with poor patient outcomes in invasive breast carcinoma

Canevari, Renata A; Marchi, Fabio A; Domingues, Maria A C; de Andrade, Victor Piana; Caldeira, José R F; Verjovski-Almeida, Sergio; Rogatto, Silvia R; Reis, Eduardo M

Published in:
Tumor Biology

DOI:
10.1007/s13277-016-5133-8

Publication date:
2016

Document version
Peer reviewed version

IDENTIFICATION OF NOVEL BIOMARKERS ASSOCIATED WITH POOR PATIENT OUTCOME IN INVASIVE BREAST CARCINOMA

Renata A. Canevari[a], Fabio A. Marchi[b], Maria A.C. Domingues[c], Victor Piana de Andrade[d], José R.F. Caldeira[e], Sergio Verjovski-Almeida[f,g], Silvia R. Rogatto[b,h*], Eduardo M. Reis[f*]

* Both authors contributed equally

[a] Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba, 12244-000, São José dos Campos, SP, Brazil.

[b] CIPE - AC Camargo Cancer Center, 01508-010, São Paulo, SP, Brazil.

[c] Departamento de Patologia, Faculdade de Medicina, Universidade do Estado de São Paulo - UNESP, 18618-000, Botucatu, SP, Brazil.

[d] Departamento de Patologia, AC Camargo Cancer Center, 01509-900, São Paulo, SP, Brazil.

[e] Departamento de Senologia, Hospital Amaral Carvalho, 17210-080, Jaú, SP, Brazil.

[f] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo - USP, 05508-900, São Paulo, SP, Brazil.

[g] Instituto Butantan, 05503-900 São Paulo, SP, Brazil.

[h], Department of Clinical Genetics, Vejle Sygehus, Vejle, DK, Institute of Regional Health, University of Southern Denmark, DK and AC Camargo Cancer Center, 01508-010, São Paulo, SP, Brazil.

***Keywords**:* breast cancer; gene expression; oligoarray; molecular genetics; prognostic signature; quantitative real-time PCR

Corresponding authors:

Silvia Regina Rogatto
Department of Clinical Genetics, Vejle Sygehus, Vejle, and Institute of Regional Health, University of Southern Denmark
Kabbleltoft 25, 7100 Vejle – Denmark
Telephone: 0045-7940 5000
E-mail: silvia.regina.rogatto@rsyd or silvia.rogatto2@gmail.com

Eduardo M. Reis
Universidade de São Paulo, Instituto de Química.
Av. Prof. Lineu Prestes, 748, Cidade Universitaria
05508-900 - São Paulo – SP, Brazil
Telephone: +55-11-3091-2173
E-mail: emreis@iq.usp.br

Abstract

Breast carcinoma (BC) corresponds to 23 % of all cancers in women, with 1.38 million new cases and 460,000 deaths worldwide annually. Despite the significant advances in the identification of molecular markers and different modalities of treatment for primary BC, the ability to predict its metastatic behavior is still limited. The purpose of this study was to identify novel molecular markers associated with distinct clinical outcomes in a Brazilian cohort of BC patients. We generated global gene expression profiles using tumor samples from 24 patients with invasive ductal BC who were followed for at least 5 years, including a group of 15 patients with favorable outcomes and another with nine patients who developed metastasis. We identified a set of 58 differentially expressed genes ($p \leq 0.01$) between the two groups. The prognostic value of this metastasis signature was corroborated by its ability to stratify independent BC patient datasets according to disease-free survival and overall survival. The upregulation of *B3GNT7, PPM1D, TNKS2, PHB*, and *GTSE1* in patients with poor outcomes was confirmed by quantitative reverse transcription polymerase chain reaction (RT-qPCR) in an independent sample of patients with BC (47 with good outcomes and eight that presented metastasis). The expression of BCL2-associated agonist of cell death (BAD) protein was determined in 1276 BC tissue samples by immunohistochemistry and was consistent with the reduced *BAD* mRNA expression levels in metastatic cases, as observed in the oligoarray data. These findings point to novel prognostic markers that can distinguish breast carcinomas with metastatic potential from those with favorable outcomes.

Key words: breast cancer; prognostic gene signature; metastasis; gene expression; tumor biomarkers

Introduction

Breast cancer (BC) is a complex and heterogeneous disease, with distinct clinical presentations in different patient and ethnic populations. Even patients affected by tumors with similar histomorphological appearances may present divergent clinical courses [1, 2]. During the last few decades, several clinical and pathological indicators, such as histological grade, tumor size, and lymph node involvement, have been used for prognostic prediction in breast cancer patients [3]. In addition, the predictive and prognostic markers applied to the management of patients include the estrogen (ESR1) and progesterone receptor (PR) statuses and human epidermal growth factor receptor 2 (Her2) amplification and/or overexpression. The clinico-pathological risk assessment is routinely based on these factors and is used to guide decisions on adjuvant systemic treatment [4]. Despite the improvements in risk stratification, the current prognostic factors show moderate accuracy in classifying breast tumors according to their clinical behaviors. It is still a clinical challenge to identify the patients who are at a low risk of relapse and have been submitted to overtreatment after optimal locoregional treatment as well as those who are at a high risk of relapse and would benefit from a more aggressive adjuvant systemic therapy and closer follow- up.

New biomarker development is required to assist clinicians in BC detection and diagnosis, risk stratification, disease subtyping, prediction of treatment responses, and surveillance, allowing personalized cancer management. The integration between novel biomarkers and routinely tested clinico-pathological features, such as hormone receptor (HR) and Her2 statuses, may guide clinicians in their systemic therapy decisions regarding both primary and metastatic tumors [5–8].

In recent years, gene expression prognostic tests for BC that are better predictors of clinical outcomes than traditional pathological factors have been developed [9–12]. One of the first commercially available and US Food and Drug Administration (FDA)-approved signatures was the 70-gene MammaPrint assay® (Agendia Inc.). MammaPrint is among the most validated multigene prognostic signatures and is widely used to stratify lymph node-negative patients as having low or high risks of distant metastases at 5 years from surgery [13]. The 21-gene Oncotype Dx® assay (Genomic Health Inc.) has also been extensively used to estimate the risk of relapse in estrogen receptor (ER)+, node-negative BC, and their chemosensitivities. A common feature between both signatures is the ability to better estimate the risk of recurrence compared to

conventional clinical criteria, such as the St. Gallen and NIH consensus criteria or web-based decision tools (e.g., Adjuvant! Online; https://www.adjuvantonline.com). This improvement in risk estimation can help to reduce the number of over treated women, which emphasizes the advantage of these gene signatures over the clinical guidelines [14, 15]. Other multigene expression-based prognostic tests interrogating different sets of genes have been made commercially available more recently, with comparable prognostic performances in BC [8]. PAM50, which is based on a 50-gene set, generates a numerical score (risk of recurrence, or ROR) that, along with the clinical features, estimates the risk of relapse at 10 years in postmenopausal women with stage I/II node-negative or stage II node-positive (one to three positive lymph nodes) and HR-positive BC [16]. The 97-gene MapQuant Dx® (Ipsogen-Qiagen) has been used to resolve intermediate grade 2 BC tumors into those with a good prognosis, grade 1-like behavior, or more aggressive grade 3-like tumors [17]. Finally, the breast cancer index (BCI; Biotheranostics, Inc.) combines the information from the expression ratio of two estrogen-regulated genes (HOXB13 :IL17BR) with the expression profile of five other genes known as the molecular grade index (MGI) to estimate the individualized risk of the late distant recurrence of breast cancer [18].

These so-called first generation multigene prognostic signatures [19] show little overlap in their gene lists, which indicates that the results from the studies are unstable [20, 21]. In fact, a single dataset may generate several different signatures with clinically relevant subgroups of breast cancer [22]. On the other hand, comparative studies and meta-analyses have demonstrated that BC prognostic gene signatures tend to have similar performances and show relatively good concordance with their prognostic ability to identify patients with worse prognoses [23–27] and could provide a refined estimate of disease-free survival, with added value beyond the current clinical indicators [28]. Despite their demonstrated efficacy, the lack of consensus and low overlap across the genes identified in these studies indicate that the potential for identifying novel genes and expression signatures that are correlated with patient outcomes in BC has not been fully exploited.

To our knowledge, there are no descriptive studies of the molecular signatures of breast carcinomas from Brazilian patients. Brazilians form one of the most heterogeneous populations in the world, as a result of five centuries of interethnic crosses between people from three continents. For this reason, the molecular signatures of breast carcinomas from Brazilian patients may lead to the identification of new molecular markers that are common among several ethnicities. Thus, the purpose of this study was to identify new

molecular markers of BC that are common in the Brazilian population and determine whether these genes can predict the patients' clinical outcomes. Global gene expression profiles generated from 24 locally invasive ductal breast carcinoma samples from patients staged M0 at diagnosis with subsequent metastasis (n = 9) and patients who remained disease-free at a minimum follow-up of 60 months (n = 15) were used to identify a 58-gene set associated with patient outcome. The robustness of this signature to predict disease outcome (disease-free survival and overall survival) was corroborated by the analysis of publicly available BC gene expression datasets. Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was used to confirm the differential expression of a selected set of candidate markers in an independent validation set of 55 new cases (comprising 47 patients with 5-year metastasis-free BC and eight patients with metastatic disease). Furthermore, the potential of the BCL-2-associated agonist of cell death promoter (BAD) protein as a new prognostic biomarker of BC was validated using a tissue microarray comprising 1276 BC samples.

Material and Methods

Patients

The study comprised 79 female patients with infiltrating ductal breast cancer with a mean age of 58±16 years (ranged from 24-94 years). The mean follow-up after the surgery was 87.5 ± 17.4 months (ranged from 57 to 117 months). The criteria for patient inclusion were no previous or concomitant diagnosis of any cancer or metastasis at diagnosis. Patients were treated with segmental resection or mastectomy, including dissection of the axillary lymph nodes, followed by radiotherapy and adjuvant systemic therapy, if indicated. None of them received radiotherapy or chemotherapy prior to surgery. After surgery, 64 patients (81%) received radiotherapy and 43 (54%) were treated for 60 months with tamoxifen (20 mg/day) at the end of chemotherapy. All patients that developed metastasis (17 cases) presented operable stage II and III breast cancers, with larger tumors (> 2 cm) often poorly differentiated, and all except one showed positive axillary lymph nodes. Of this group, the majority (13 patients) received chemotherapy following surgery. From 14 patients with positive ER status, the majority (11 patients) was also treated with hormone therapy.

Fresh tissue samples were macro-dissected (>80% of tumor cells) and used for RNA isolation and subsequent gene expression analysis. Thirty-eight cases were used for global gene expression analysis using oligoarrays and 55 cases used for qRT-PCR analysis of candidate genes. In addition, tumor tissue samples from 1,276 ductal breast carcinomas were evaluated for protein expression with immunohistochemistry. The samples were obtained from Amaral Carvalho Hospital, Jau (SP, Brazil) and AC Camargo Cancer Center (SP-Brazil). Written informed consent from all patients was obtained during the collect period, and the study was reviewed and approved by the Ethics Committees from both Institutions (CEP FHAC 340/04 and CEP ACCC 1155/08).

Analysis design

The expression profiles from tumor samples obtained from 38 patients were used to identify the molecular signatures that were correlated with the Her2 and Ki67 expression statuses, two of the currently used prognostic markers. The detailed clinico-pathological data from these patients is shown in Supplementary Table 1. A subset of 24 cases (test set) with known outcomes was used to identify a gene expression

signature associated with the development of metastasis. The ability of the "metastasis signature" to stratify patients based on the time to develop metastasis was compared to a combined prognostic score based on traditional clinico-pathological criteria that assigned patients to a "good outcome" group comprised of 15 patients who remained free of disease at least 5 years after surgery. Nine patients who developed metastasis were assigned to the "poor outcome" group. The criteria used to assign a patient to the good outcome group were the presence of at least five of the following parameters: (i) tumor clinical stage I or IIA; (ii) tumor size less than 2 cm; (iii) histopathological grade I or II; (iv) negative axillary lymph node status or less than four nodes compromised; (v) ER- positive (scores 2+/3+) in the immunohistochemistry assay; (vi) Her2-negative or (1+) in the immunohistochemistry assay; and (vii) low proliferation index (<25 % of cells with Ki67-positive staining). To verify if the development of metastasis in this sample set could be explained by the traditional clinico-pathological factors alone, the association between these variables was evaluated using Fisher's exact test (GraphPad Prism Statistical Software version 5.0, San Diego, CA, USA) and statistical significance was designated at $p \leq 0.05$ (Supplementary Table 2). The histological grade, proliferation index, and adjuvant chemotherapy were individual characteristics that were statistically related to the development of metastasis in the group of patients analyzed.

The independent validation set (55 samples) included 47 samples from patients with good outcomes and eight samples from patients with poor outcomes. A summary of the clinico-pathological data from the 79 patients with a known follow-up that comprised the samples used in the gene expression experiments (test + validation set) is shown in Supplementary Table 3. Information on the additional 1276 cases used in tissue microarray validation experiments is provided in Supplementary Table 4.

RNA isolation and global gene expression analysis

The total RNA was extracted using the RNeasy mini Kit (Qiagen GmbH, Germany). RNA samples were treated with DNAse I Amplification Grade (Life Technologies, Rockville, MD, USA) according to the manufacturer's recommendations. The quantification and integrity assessment of RNA samples were performed using the ND-1000 spectrophotometer (NanoDrop Inc.) and Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. Only samples with RIN (RNA integrity number) ≥ 7 were considered further. Gene expression profiles were collected using Whole Human Genome CodeLink bioarrays (GE Healthcare

Buckinghamshire, UK) that interrogate approximately 54,000 transcripts (GEO accession GPL11010). In brief, first-strand cDNA was produced using 2 μg of total RNA from each sample, Superscript II reverse transcriptase, and a T7-poly-dT primer. Second-strand cDNA was produced using RNase H and *E. coli* DNA polymerase I. Double-stranded cDNA was column purified (QIAquick, Qiagen) and biotin-labeled cRNA targets were generated by an *in vitro* transcription reaction using T7 RNA polymerase and biotin-11-UTP (Perkin Elmer-Foster City, CA, USA). Fragmented cRNA from each sample was hybridized to CodeLink microarrays overnight at 37 °C in a shaking incubator at 300 rpm. After post-hybridization washes, hybridized targets were revealed by incubating the arrays with a Cy5-Streptavidin conjugate. All reagents used for the synthesis and fragmentation of cRNA were provided in the CodeLink expression assay kit (GE Healthcare). Signal of the Cy5-dye from hybridized targets were detected with an arrayWoRx Biochip Reader (Applied Precision LLC, Issaquah, WA, USA). CodeLink Expression Analysis software (GE Healthcare) was used to extract background-subtracted spot intensities from microarray images. A set of 23,566 probes that showed valid measurements in at least 60% of the samples was further analyzed. Intensity data were normalized across samples using the quantile method [29]. Raw and normalized expression measurements are deposited at the Gene Expression Omnibus (GEO) under accession number GSE73383.

A Signal-to-Noise Ratio (SNR) statistical test [30] was used to identify gene expression signatures correlated with pathological and clinical parameters of samples. The statistical significance of the differential expression (p-values) was ascertained by bootstrap (1,000 random permutations). Expression profiles of transcripts selected in each analysis were clustered hierarchically (UPGMA with Euclidean distance) and visualized using the Spotfire Decision Site software (TIBCO Spotfire, Somerville, MA, USA). The robustness of the gene expression signature of metastasis was evaluated by leave-one-out resampling [11] considering a $p \leq 0.01$ cutoff.

Reverse Transcription Quantitative Polymerase Chain Reaction (qRT-PCR)

Real time quantitative RT-PCR amplifications for a subset of genes selected from the metastasis signature (*B3GNT7, PPM1D, TNKS2, PHB,* and *GTSE1* genes) were performed in duplicate on an ABI Prism 7000

Sequence Detection System (Applied Biosystems, Foster City, CA, USA), using Power SYBR Green I reagent (Applied Biosystems). These genes were chosen by their importance in cell proliferation or their association with biological processes related to metastasis. The primers shown in Supplementary Table 5 were designed using the Primer Express software (v3.0; Applied Biosystems). The standard curves of the targets and reference genes showed similar amplification efficiencies (> 90%). The qPCR amplification data were analyzed using the Sequence Detection System software (v1.0; Applied Biosystems). Only replicates with low variability (i.e. Δ cycle quantification < 0.5) were considered for further analyses. *GAPDH* was selected as the reference gene [31]. The relative expression of target genes was calculated according to the ΔCt method [32].

The normalized expression levels of target genes in the tumor samples were represented as fold-change relative to their abundance in a pool of non-tumor breast tissues (Relative Quantification, RQ), calculated as follows: $2^{-(\Delta Ct \text{ test sample} - \Delta Ct \text{ control sample})}$. The statistical analysis of qRT-PCR results was performed using GraphPad InStat software (version 3.00). The Mann-Whitney test was used to ascertain the statistical significance of the expression levels evaluated by RT-qPCR. The correlation between the oligoarray and RT-qPCR data was evaluated using Spearman's correlation test. For this analysis, the raw expression values estimated by each of the two methodologies were used, and the statistical significance was ascertained using a $p < 0.05$ threshold.

In silico validation and functional analysis of the metastasis signature

The individual abilities of the genes present in the metastasis signature to classify patients according to their disease outcomes (presence/absence of metastasis) were evaluated by receiver operating characteristic (ROC) curves using GraphPad Prism 7 software (GraphPad Software, Inc.). For each gene, the ability to correctly identify disease-free vs. metastatic patients was evaluated by calculating the area under the ROC curve (AUC) using a non-parametric method. The ability of the 58 signatures to predict patient outcome as a multigene biomarker was tested using SurvExpress, a web-based tool that provides survival analysis and risk assessment publicly available cancer datasets [33]. Briefly, the tool applies the Cox proportional hazard regression model to generate risk scores (also called prognostic indexes) that relate patient survival (disease-free or overall survival) to the expression of a given gene list. The high- and low-risk patient groups were

generated by splitting the samples at the median after ranking them by their risk scores. The log-rank test was used to ascertain the statistical significance of the difference be- tween the survival curves.

Ingenuity Pathway Analysis (IPA v8.0, Ingenuity® Systems, Redwood City, CA, USA; http://www.ingenuity. com) was used to identify the enriched pathways and biological interaction networks of the differentially expressed genes identified in the oligoarray analysis. Fischer's exact test was applied to identify the significant networks and pathways that were represented within the respective gene sets. In parallel, we searched for significantly enriched pathways among the metastasis signature with the software KOBAS 2.0 [34], which incorporates knowledge across 1327 species from five pathway databases (KEGG PATHWAY, PID, BioCyc, Reactome, and PANTHER) and five human disease databases (OMIM, KEGG DISEASE, FunDO, GAD, and NHGRI GWAS Catalog). Only pathways identified simultaneously by IPA and KOBAS with a $p \leq 0.05$ were considered for further interpretation (Table 1).

Protein Expression by Immunohistochemistry using Tissue Microarrays (TMA)

Tissue microarrays (TMAs) with core biopsies from 1276 ductal breast carcinomas were assembled as previously described using a Manual Tissue Microarrayer (Beecher Instruments®, Silver Springs, USA) [35]. Tissue cores with a dimension of 1.0 mm from each specimen were punched and arrayed in quadruplicate on recipient paraffin blocks. Paraffin-embedded breast tumor samples were sectioned (3 μm) and mounted on silane-coated glass slides for hematoxylin-eosin (HE) staining and immunohistochemistry (IHC).

IHC was performed using the anti-BAD (clone Y208, Epitomics diluted in 1:1500, antigen retrieval in EDTA/Tris, pH 9.0) primary antibody and secondary antibodies (Advanced TM HRP Link, Dako Cytomation, K0690, Denmark). Positive (BAD, prostate tissue) and two negative controls (omitting the primary antibody and replacing the primary antibody with normal rabbit serum) were assessed by IHC. BAD expression in each sample core was scored as positive or negative following a visual inspection of the cytoplasmic immunostaining. The final scores (median of the four cores) were obtained based on the immunostaining intensity in the cytoplasm and were denominated as "negative/weak" (score 0–1) or "positive" (score 2–3). A chi-square test was applied to determine the association strength between the

categorical variables ($p < 0.05$). Univariate Kaplan-Meier (KM) survival curves were calculated for systemic disease-free survival (SDFS) and cancer-specific survival (CSS) probabilities by considering a minimum 60-month follow-up after surgery. The log-rank test was used to assess the statistical significance of the KM curves. IHC analysis was blinded to the outcomes and clinical aspects of each tumor specimen. A minimum 5-year follow-up after surgery was used for both SDFS and CSS. Multivariate analysis was performed using the Cox proportional hazard model. The statistical analyses were performed using MedCalc v. 15.11.0 (MedCalc Software, Belgium), GraphPad Prism3 (San Diego, CA, USA), and SPSS version 15.0 (SPSS, Chicago, IL, USA).

Results

Gene expression signature associated with metastatic ductal breast carcinomas

First, the global gene expression profiles were evaluated in 38 BC samples to identify gene signatures that were correlated with the commonly used prognostic markers of breast cancer outcome (expression of ESR1/PR, Ki67, and Her2), with the aim of distinguishing patients who developed metastasis from those that remained disease-free. Using a two-class supervised analysis based on a SNR with permutation [30], we identified statistically significant gene expression signatures that were correlated with the expression of estrogen/progesterone recep- tors ($p \leq 0.005$), Ki67 ($p \leq 0.001$), or Her2 ($p \leq 0.001$). Although these expression signatures accurately discriminate samples based on each prognostic marker, they are poor estimators of risk of metastatic progression in patients with invasive BC (Supplementary Figure 1).

Next, we sought to identify a gene expression signature that was correlated with disease outcome, i.e., the development of metastasis. For this analysis, we only included samples from the 15 patients with a good prognosis that remained free of disease for at the least 5 years. SNR with permutation was used and identified 200 full-length or partial transcript (EST) sequences that were differentially expressed ($p \leq 0.01$) between the samples from patients who developed metastasis (nine cases) and those from patients with a good prognosis that remained disease-free (15 cases). Hierarchical clustering of this transcript set correctly grouped the samples according to their metastatic outcome, with the exception of one sample (Supplementary Figure 2 and Supplementary Table 6).

To identify the most robust markers of metastasis, a leave- one-out (LOO) cross-validation procedure [11] reduced the initial 200-gene signature to a 58-transcript set comprising sequences that are present in all LOO datasets at a threshold of $p \leq 0.01$ (Fig. 1). A classifier based on the expression of the 58-transcript set was able to correctly discriminate all samples according to patient outcome (≥5-year disease-free versus metastasis), with the exception of one case (P598) (Fig. 1, right panel).

To evaluate the potential of this 58-transcript set to discriminate breast cancer patients according to the time required to develop metastasis, we generated a KM survival curve and this analysis showed significant discriminating power ($p < 0.0001$) in stratifying the patient samples according to time to develop metastasis (Fig. 2a). For comparison, survival curves were generated based on Her2 and Ki67 expression statuses or the tumor histopathological grade, three currently used parameters to infer breast cancer outcome. It is apparent that the 58-gene set outperformed Her2 ($p = 0.199$) and Ki67 ($p = 0.115$) (Fig. 2b, c, respectively) and was comparable to the histopathological grade ($p = 0.0082$) (Fig. 2d) in discriminating patients according to their risk of developing metastasis. ROC curves were constructed to assess the individual potential of each gene from the metastasis signature to classify patients according to their disease outcomes. For each gene, the AUC was estimated to measure the sensitivity and specificity of the biomarker to correctly select the disease-free and metastatic cases. The AUC values ranged from 0.933 to 0.807 (Supplementary Table 7), suggesting good biomarker performance.

Biological network analysis

Enrichment analysis using two methods (IPA and KOBAS) identified pathways that are significantly over-represented ($p < 0.05$) among the 58-gene metastasis signature (Table 1). These included TRAF6 mediated NF-kB activation, p53 pathway, Toll Like Receptor cascades (TLR5, TLR7/8, TLR6/2; TLR3/TLR4; TLR3; TLR2, TLR9 and TLR10), TRAF6 mediated induction of NFkB and MAP kinases; Eph-ephrin signaling; PECAM1 interactions; HDL-mediated lipid transport; p75NTR signaling complexes; IRAK1 recruitment of the IKK complex; WNT5A-dependent internalization of FZD4 and signaling by Rho GTPases (Table 1).

Validation of molecular markers of metastasis in breast cancer

To document the expression of the genes identified in the metastasis signature in independent datasets, the 58-gene signature was cross-referenced with the list of genes differentially expressed in metastatic breast

cancer studies retrieved from the data collected and processed at the NCBI-GEO database (http://www.ncbi.nlm.nih.gov/geo/) originated from ten studies reported in the literature [11, 36-45] . The cut-off values for selecting gene lists from each analysis were adjusted between $p < 0.001$ to $p < 0.05$ to retrieve a comparable number of genes in each analysis. In this meta-analysis, 12 transcripts were found in common. From this set, 8 showed the same direction of transcriptional change observed in our study in 50% or more of the studies used in the meta-analysis (Table 2). Of note, only the *GTSE1* and *ZNF664* genes from our 58-gene signature showed overlap with the genes present in the commercially available metastasis signature MamaPrint® [11] and none of them were found in common with the Oncotype DX qRT-PCR based signature [46].

Additionally, the ability of the 58-gene signature to predict patient outcome was evaluated in different publicly available breast cancer datasets [42, 47]. Using the SurvExpress multigene biomarker validation tool [33], we found that the metastasis signature was very effective to discriminate patients at higher risk of developing metastasis in the breast cancer datasets described by van de Vijver et al. [42] (n = 249, $p = 4.1e^{-11}$, HR= 5.2, 95% CI 2.9 − 9.5; Fig. 3a) [41] and by Kao et al. [47] (n = 367, $p = 7.9\ e^{-8}$, HR= 3.6, 95% CI 2.2 – 6.0; Fig. 3b). The metastasis signature also stratified these patients according to overall survival ($p = 1.8\ e^{-9}$, HR= 4.5, 95% CI 2.6 – 7.7 and $p = 6.7\ e^{-10}$, HR= 4.6, 95% CI 2.7 – 7.9, respectively) (Supplementary Fig 3a,b). The potential of the 58-metastasis signature to predict disease outcome was further corroborated by its ability to discriminate patients according to the overall survival using RNAseq gene expression data generated by the TCGA consortium from a cohort of 502 BC patients ($p = 3.2\ e^{-9}$, HR= 5.2, 95% CI 2.9 – 9.5; Supplementary Figure 3, panel c) [45].

We focused on a subset of genes from the metastasis signature (*B3GNT7, PPM1D, TNKS2, PHB,* and *GTSE1*) due to their importance in cell proliferation or because they are directly or indirectly involved in biological processes related to metastasis**.** Consistent with oligoarray data, all five genes were significantly overexpressed ($p < 0.05$) in tumor samples from patients that developed metastasis relative to those that remained disease-free when the same samples (24-samples test set) were evaluated by RT-qPCR (Supplementary Fig. 4, left panels). More importantly, similar results were observed in an independent set of samples (validation set) comprising 47 primary tumor samples from patients that remained disease free after

surgery, and 8 primary tumor samples from patients that developed metastasis (Supplementary Fig. 4, right panels).

The BCL2-associated agonist of cell death (*BAD*) gene was downregulated in tumors from patients that developed metastasis (Supplementary Table 6). BAD protein expression was evaluated in a TMA comprising 1276 primary BC tissue samples from patients who developed metastasis (432 samples) or remained disease-free (370 samples) (Table 3). The tumors from patients who developed distant metastasis during the follow-up period exhibited lower expression levels of BAD ($p < 0.0001$). Reduced BAD expression was also statistically associated with positive lymph node status ($p = 0.0035$), negative ER ($p < 0.0001$), negative PR ($p < 0.0001$), negative Her2 ($p = 0.0182$), more advanced pathological stages ($p = 0.0002$), and tumor size (T3-T4) ($p = 0.0002$). Interestingly, the percentage of cases with absent/low BAD expression (score 0) increased according to the number of lymph nodes involved. No statistically significant association was observed between BAD protein expression and age, family history, presence of distant metastasis at diagnosis, and histological grade (Table 3). In addition, the survival analysis showed that downregulation of BAD in primary tumors was significantly associated with SDFS ($p = 0.0001$, HR = 1,63; 95% CI 1,23 to 2,15) and CSS ($p < 0.0001$, HR = 1,65; 95% CI 1,25 to 2,18) (Fig. 4). However, in the multivariate analyses BAD expression was not statistically associated with SDFS or CSS (data not shown).

Discussion

In this study, a 58-gene expression signature associated with the clinical outcomes of BC in a cohort of Brazilian patients was identified, thus revealing new molecular markers with prognostic potential. The cross-referencing of this 58-gene set with the breast cancer prognosis signatures from 10 studies reported in the literature [11, 36–44] revealed 10 genes with concordant gene expression changes in at least one other analysis. Despite the discrepancies, this analysis reinforces the prognostic value of some genes, such as *ITGAV, GTSE1*, and *AP2B*, in breast cancer. Disagreement between gene signatures is frequently observed and may originate from individual variations, the molecular and clinical heterogeneity of tumors, the use of different platforms, and differences in patient selection, data normalization, methods of analysis, and other experimental choices [48]. In this regard, we observed a modest overlap with the genes present in the MammaPrint® signature and no overlap with the genes of the Oncotype Dx® signature.

The 58-gene metastasis signature was further evaluated in three independent BC patient datasets with follow-up infor- mation. The metastasis gene set was able to stratify BC metastasis signature for risk stratification and highlight the value of this gene set as a resource to select candidates for biomarker development in BC.

The 58-gene set identified in this study was enriched in molecular pathways that may contribute to the metastatic dissemination of breast tumors. Noteworthy, the toll-like receptor (TLR) cascade pathway was represented by nine members of the TLR family (*TLR5, TLR7/8, TLR6/2; TLR3/TLR4; TLR3; TLR2, TLR9*, and *TLR10*). TLRs initiate a series of downstream signaling events that drive cellular responses, including the production of cytokines, chemokines, and other inflammatory mediators. In addition to driving inflammatory responses, TLRs also regulate cell proliferation and survival, which serves to expand useful immune cells and integrate inflammatory responses and tissue repair processes [49]. In colorectal cancer, TLR signaling may promote metastasis by activating the expression of integrins and chemokine receptors that facilitate tumor cell migration and the colonization of distal sites [50]. Several reports suggest that the TLR signaling pathway may play a supporting role in the secretion of pro-inflammatory cytokines/chemokines, aggressive tumor behavior (e.g., NF-κB activity), cell proliferation, cell invasion, cell migration, and metastasis in breast tumor cells [51]. Our study indicates that augmented TLR signaling is also a feature of metastatic breast tumors.

Genes associated with the p53 pathway were also overrepresented among the metastasis-associated gene set. The p53 protein directly controls the transcription of genes that are involved in canonical metastasis pathways, including cell adhesion, motility, invasion, EMT, stemness, and ECM interactions [52]. Eph-ephrin signaling genes were also identified in this signature. This family of molecules is involved in many aspects of both normal and carcinogenic developmental processes. In breast cancer, the roles of Eph receptors are extremely versatile and multifaceted, as they either promote or suppress tumor functions. The tumor-promoting effects seem to prevail [53], which is in agreement with our findings.

Molecular pathways associated with BC invasiveness have also been identified. TRIP10, also known as Cdc42- interacting protein-4 (CIP4), was found upregulated in the tumors from patients who developed metastasis. Higher CIP4 levels have been shown to be significantly associated with a greater risk of metastatic progression in triple-negative breast cancer patients [54]. In addition to its value as a biomarker

for patient outcome, it has been demonstrated that CIP4 localizes to cell invadopodia and positively contributes to cell invasion and metastasis in in vivo and in vitro breast cancer models, pointing to the functional relevance of this gene for metastatic dissemination in BC [54, 55].

B3GNT7, PPM1D, TNKS2, PHB and *GTSE1* genes were significantly differentially expressed in the independent validation set. These genes were upregulated in the patients that developed metastasis in comparison with those that remained disease-free. *B3GNT7* and *PPM1D* genes were previously associated with poor outcome of breast cancer patients [56]; however the value of *TNKS2, PHB,* and *GTSE1* genes as prognostic markers is currently unknown. *B3GNT7* encodes an enzyme that is involved in a broad variety of biological functions that are mainly related to cell-cell communication [57], differentiation [58], and certain infectious diseases [59] besides promoting cell motility and invasion in vitro [60]. The increased expression of *B3GNT7* in metastatic cases, compared to cases without metastasis, indicates that this gene may be an important prognostic marker in breast cancer.

PPM1D is a hotspot for gene amplification in breast cancer [61] and has been detected in both breast cancer cell lines and primary breast tumors, which suggests its involvement in cancer development [56, 62]. *PPM1D* is involved in the regulation of several essential signaling pathways that are implicated in BC pathogenesis, such as the inhibition of tumor suppressor activities of TP53 [63], p16INK4A, and p19ARF [64] and the activation of oncogenes, such as *RAS, MYC*, and *NEU*, thus promoting cellular transformation [65, 66]. In addition, PPM1D targets other key stress response kinases, such as *ATM, CHK1, CHK2*, and *UNG2*, which function in the DNA damage response and repair [63, 67]. These findings indicate that *PPM1D* is involved in the regulation of several essential signaling pathways that are implicated in breast cancer pathogenesis, and, therefore, it could be considered as a potential therapeutic target in BC patients.

TNKS2 (codified enzyme tankyrase, TRF1-interacting an- kyrin-related ADP-ribose polymerase 2) positively regulates telomere length, and its expression has been detected in meningiomas and breast cancer [68]. The tumor suppressor gene *PHB* (prohibitin) encodes an anti-proliferative protein that functions by interacting with the Rb protein and its family members. Their putative association with the immunoglobulin (Ig)M receptor and estrogen receptor has suggested to have a possible role in cellular signaling, whereas the binding of prohibitin to E2F proteins has been taken as an indication of a possible role of PHB in

transcriptional regulation during the cell cycle [69]. *PHB* is overexpressed in several carcinomas, and its role in tumorigenesis has been associated with the modulation of cell cycle control via Ras-Raf signaling, cell migration, mitochondrial physiology, and apoptosis [70, 71]. In this study, *TNKS2* and *PHB* were upregulated in cases that presented metastasis in comparison to cases without metastasis in both the test and validation sets. Based on these results, we suggest that these genes may be potential prognostic markers of breast cancer.

GTSE1, which is present in the 58-gene metastasis signature, was also differentially expressed in patients with metastatic breast cancer, with the same signal direction as the TGCA analysis (2015). This gene (G2 and S phase- expressed-1) plays a role in the G2 phase of the mitotic cell cycle. Its protein is co-localized with tubulin, suggesting that GTSE1 is associated with microtubules [72]. A physical interaction occurs between the C-terminal region of GTSE1 and the C-terminal regulatory domain of p53, which is necessary and sufficient to downregulate p53 activity. After DNA damage, GTSE1 could play a dual role during the G2 checkpoint, promoting a delay of the G2 to M transition and, at the same time, protecting these cells from p53-dependent apoptosis [73]. In one study, GTSE1 was found to be highly expressed in oral tongue squamous cell carcinoma samples and significantly associated with patients presenting lymph node metastasis [74]. To our knowledge, there are no previous studies associating the clinico-pathological significance of *GTSE1* in patients with breast carcinoma.

We observed reduced expression of BAD (BCL2-associated agonist of cell death) in the tumors of patients that developed metastasis compared to patients that remained metastasis-free. We found significant associations between BAD downregulation and occurrence of distant metastasis during the ≥ 5-year follow-up period ($p<0.0001$), positive nodal status ($p=0.0035$), negative ER ($p<0.0001$), negative PGR ($p<0.0001$), negative Her2 ($p=0.0182$) and tumor size (T3-T4) ($p=0.0002$). In addition, the survival analysis showed that reduced BAD expression was statistically associated with shorter systemic disease free-survival ($p=0.001$) and cancer specific survival ($p<0.001$). Cekanova et al. [75] suggest that, in addition to the effect on apoptosis, *BAD* conveys anti-metastatic effects and is a valuable prognostic marker in breast cancer. In line with this study, the gene and protein expression levels that were downregulated in our 58-gene set and protein analysis suggest that expression level of *BAD* is a clinically relevant prognostic marker in breast carcinomas.

In conclusion, in this study we identified a 58-gene expression signature able to distinguish patients according to their risk of developing metastasis or overall survival more effectively than clinico-pathological factors currently used for BC prognosis. This metastasis signature showed prognostic value in independent BC patient datasets. A subset of these genes was validated by transcript or protein expression analyses in an independent panel of samples, and the results highlighted candidates with the potential to stratify patients according to their risk of developing distant metastasis after diagnosis. Further studies with extended panels of patient samples are warranted to validate the clinical relevance of the novel biomarker candidates identified here.

Acknowledgments

This work was mainly supported by a grant from the Conselho Nacional de Desenvolvimento Científico e Tecnologico (CNPq) (Edital MCT/CNPq/CT-Biotecnologia n° 010/2004). Additional funding was provided by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). E.M.R. and S.V.A. received investigator fellow- ship awards from CNPq. The authors thank Sandra Dringo Linde for her expert technical assistance during this study.

Authors' contributions

SVA, SRR, and EMR conceived and designed the experiments. RAC and MACD performed the experiments. EMR, FM, and RAC analyzed data. JRFC and VPA contributed samples. RAC, FM, VPA, SRR, and EMR drafted or revised the manuscript. All authors read and approved the final manuscript.

Compliance with ethical standards

Written informed consent was obtained from all patients during the collection period, and the study was reviewed and approved by the Ethics Committees from both institu- tions (CEP FHAC 340/04 and CEP ACCC 1155/08).

Conflicts of interest

None

References

1 Veta M, Pluim JP, van Diest PJ, Viergever MA: Breast cancer histopathology image analysis: A review. IEEE Trans Biomed Eng 2014;61:1400-1411.

2 Dumalaon-Canaria JA, Hutchinson AD, Prichard I, Wilson C: What causes breast cancer? A systematic review of causal attributions among breast cancer survivors and how these compare to expert-endorsed risk factors. Cancer Causes Control 2014;25:771-785.

3 Colombo PE, Milanezi F, Weigelt B, Reis-Filho JS: Microarrays in the 2010s: The contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. Breast Cancer Res 2011;13:212.

4 Rosa M. Advances in the molecular analysis of breast cancer: path- way toward personalized medicine. Cancer Control. 2015;22:211–9.

5 Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346-352.

6 Reis-Filho JS, Weigelt B, Fumagalli D, Sotiriou C: Molecular profiling: Moving away from tumor philately. Sci Transl Med 2010;2:47ps43.

7 Taherian-Fard A, Srihari S, Ragan MA: Breast cancer classification: Linking molecular mechanisms to disease prognosis. Brief Bioinform 2015;16:461-474.

8 Toss A, Cristofanilli M: Molecular characterization and targeted therapeutic approaches in breast cancer. Breast Cancer Res 2015;17:60.

9 Patani N, Martin LA, Dowsett M: Biomarkers for the clinical management of breast cancer: International perspective. International journal of cancer Journal international du cancer 2013;133:1-13.

10 Peppercorn J, Perou CM, Carey LA: Molecular subtypes in breast cancer evaluation and management: Divide and conquer. Cancer Invest 2008;26:1-10.

11 van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-536.

12 Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-679.

13 Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J, Roumen RM, Knauer M, van 't Veer LJ, Sonke GS, Rutgers EJ, van de Vijver MJ, Linn SC: A prospective evaluation of a breast cancer prognosis signature in the observational raster study. Int J Cancer 2013;133:929-936.

14 Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thurlimann B, Senn HJ: Thresholds for therapies: Highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2009. Ann Oncol 2009;20:1319-1329.

15 Sotiriou C, Pusztai L: Gene-expression signatures in breast cancer. N Engl J Med 2009;360:790-800.

16 Filipits M, Nielsen TO, Rudas M, Greil R, Stoger H, Jakesz R, Bago-Horvath Z, Dietze O, Regitnig P, Gruber-Rossipal C, Muller-Holzner E, Singer CF, Mlineritsch B, Dubsky P, Bauernhofer T, Hubalek M, Knauer M, Trapl H, Fesl C, Schaper C, Ferree S, Liu S, Cowens JW, Gnant M: The pam50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. Clin Cancer Res 2014;20:1298-1305.

17 Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, Foekens JA, Klijn JG, Larsimont D, Buyse M, Bontempi G, Delorenzi M, Piccart MJ, Sotiriou C: Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007;25:1239-1246.

18 Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, Harris A, Goss P, Sotiriou C, Erlander M, Sgroi D: A five-gene molecular grade index and hoxb13:Il17br are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008;14:2601-2608.

19 Reis-Filho JS, Pusztai L: Gene expression profiling in breast cancer: Classification, prognostication, and prediction. Lancet 2011;378:1812-1823.

20 Ein-Dor L, Kela I, Getz G, Givol D, Domany E: Outcome signature genes in breast cancer: Is there a unique set? Bioinformatics 2005;21:171-178.
21 Gormley M, Dampier W, Ertel A, Karacali B, Tozeren A: Prediction potential of candidate biomarker sets identified and validated on gene expression data from multiple datasets. BMC Bioinformatics 2007;8:415.
22 Korkola JE, Blaveri E, DeVries S, Moore DH, 2nd, Hwang ES, Chen YY, Estep AL, Chew KL, Jensen RH, Waldman FM: Identification of a robust gene signature that predicts breast cancer outcome in independent data sets. BMC Cancer 2007;7:61.
23 Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM: Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560-569.
24 Haibe-Kains B, Desmedt C, Piette F, Buyse M, Cardoso F, Van't Veer L, Piccart M, Bontempi G, Sotiriou C: Comparison of prognostic gene expression signatures for breast cancer. BMC Genomics 2008;9:394.
25 Paik S: Is gene array testing to be considered routine now? Breast 2011;20 Suppl 3:S87-91.
26 Saini A, Hou J, Zhou W: Breast cancer prognosis risk estimation using integrated gene expression and clinical data. Biomed Res Int 2014;2014:459203.
27 Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutz F, Goldstein DR, Piccart M, Delorenzi M: Meta-analysis of gene expression profiles in breast cancer: Toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008;10:R65.
28 Gyorffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L: Multigene prognostic tests in breast cancer: Past, present, future. Breast Cancer Res 2015;17:11.
29 Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19:185–93.
30 Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999;286:531-537.
31 Hu Z, Bonifas JM, Aragon G, Kopelovich L, Liang Y, Ohta S, Israel MA, Bickers DR, Aszterbaum M, Epstein EH, Jr.: Evidence for lack of enhanced hedgehog target gene expression in common extracutaneous tumors. Cancer Res 2003;63:923-928.
32 Pfaffl MW: A new mathematical model for relative quantification in real-time rt-pcr. Nucleic Acids Res 2001;29:e45.
33 Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, Martinez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, Tamez-Pena JG, Trevino V: Survexpress: An online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 2013;8:e74250.
34 Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: Kobas 2.0: A web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 2011;39:W316-322.
35 Nagai MA, Fregnani JH, Netto MM, Brentani MM, Soares FA: Down-regulation of phlda1 gene expression is associated with breast cancer progression. Breast Cancer Res Treat 2007;106:49-56.
36 Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C: Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the transbig multicenter independent validation series. Clin Cancer Res 2007;13:3207-3214.
37 Jezequel P, Campone M, Roche H, Gouraud W, Charbonnel C, Ricolleau G, Magrangeas F, Minvielle S, Geneve J, Martin AL, Bataille R, Campion L: A 38-gene expression signature to predict metastasis risk in node-positive breast cancer after systemic adjuvant chemotherapy: A genomic substudy of pacs01 clinical trial. Breast Cancer Res Treat 2009;116:509-520.
38 Karlsson E, Delle U, Danielsson A, Olsson B, Abel F, Karlsson P, Helou K: Gene expression variation to predict 10-year survival in lymph-node-negative breast cancer. BMC Cancer 2008;8:254.
39 Minn AJ, Kang Y, Serganova I, Gupta GP, Giri DD, Doubrovin M, Ponomarev V, Gerald WL, Blasberg R, Massague J: Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. J Clin Invest 2005;115:44-55.
40 Molloy TJ, Roepman P, Naume B, van't Veer LJ: A prognostic gene expression profile that predicts circulating tumor cell presence in breast cancer patients. PLoS One 2012;7:e32426.

41 Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, Borresen-Dale AL: Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: Gene expression analyses across three different platforms. BMC Genomics 2006;7:127.
42 van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.
43 Wang DY, Done SJ, McCready DR, Boerner S, Kulkarni S, Leong WL: A new gene expression signature, the clinicomolecular triad classification, may improve prediction and prognostication of breast cancer at the time of diagnosis. Breast Cancer Res 2011;13:R92.
44 Zhao H, Langerod A, Ji Y, Nowels KW, Nesland JM, Tibshirani R, Bukholm IK, Karesen R, Botstein D, Borresen-Dale AL, Jeffrey SS: Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 2004;15:2523-2536.
45 TCGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.
46 Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-2826.
47 Kao KJ, Chang KM, Hsu HC, Huang AT: Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: Implications for treatment optimization. BMC Cancer 2011;11:143.
48 Reis-Filho JS, Westbury C, Pierga JY: The impact of expression profiling on prognostic and predictive testing in breast cancer. J Clin Pathol 2006;59:225-231.
49 Li X, Jiang S, Tapping RI: Toll-like receptor signaling in cell proliferation and survival. Cytokine 2010;49:1-9.
50 Luddy KA, Robertson-Tessi M, Tafreshi NK, Soliman H, Morse DL: The role of toll-like receptors in colorectal cancer progression: Evidence for epithelial to leucocytic transition. Front Immunol 2014;5:429.
51 Kidd LC, Rogers EN, Yeyeodu ST, Jones DZ, Kimbro KS: Contribution of toll-like receptor signaling pathways to breast tumorigenesis and treatment. Breast Cancer (Dove Med Press) 2013;5:43-51.
52 Powell E, Piwnica-Worms D, Piwnica-Worms H: Contribution of p53 to metastasis. Cancer Discov 2014;4:405-414.
53 Kaenel P, Mosimann M, Andres AC: The multifaceted roles of eph/ephrin signaling in breast cancer. Cell Adh Migr 2012;6:138-147.
54 Cerqueira OL, Truesdell P, Baldassarre T, Vilella-Arias SA, Watt K, Meens J, Chander H, Osorio CA, Soares FA, Reis EM, Craig AW: Cip4 promotes metastasis in triple-negative breast cancer and is associated with poor patient prognosis. Oncotarget 2015;6:9397-9408.
55 Pichot CS, Arvanitis C, Hartig SM, Jensen SA, Bechill J, Marzouk S, Yu J, Frost JA, Corey SJ: Cdc42-interacting protein 4 promotes breast cancer cell invasion and formation of invadopodia through activation of n-wasp. Cancer Res 2010;70:8347-8356.
56 Rauta J, Alarmo EL, Kauraniemi P, Karhu R, Kuukasjarvi T, Kallioniemi A: The serine-threonine protein phosphatase ppm1d is frequently activated through amplification in aggressive primary breast tumours. Breast Cancer Res Treat 2006;95:257-263.
57 Demetriou M, Granovsky M, Quaggin S, Dennis JW: Negative regulation of t-cell activation and autoimmunity by mgat5 n-glycosylation. Nature 2001;409:733-739.
58 Nakamura N, Yamakawa N, Sato T, Tojo H, Tachi C, Furukawa K: Differential gene expression of beta-1,4-galactosyltransferases i, ii and v during mouse brain development. J Neurochem 2001;76:29-38.
59 Teneberg S, Leonardsson I, Karlsson H, Jovall PA, Angstrom J, Danielsson D, Naslund I, Ljungh A, Wadstrom T, Karlsson KA: Lactotetraosylceramide, a novel glycosphingolipid receptor for helicobacter pylori, present in human gastric epithelium. J Biol Chem 2002;277:19709-19719.
60 Kataoka K, Huh NH: A novel beta1,3-n-acetylglucosaminyltransferase involved in invasion of cancer cells as assayed in vitro. Biochem Biophys Res Commun 2002;294:843-848.
61 Sinclair CS, Rowley M, Naderi A, Couch FJ: The 17q23 amplicon and breast cancer. Breast Cancer Res Treat 2003;78:313-322.
62 Li J, Yang Y, Peng Y, Austin RJ, van Eyndhoven WG, Nguyen KC, Gabriele T, McCurrach ME, Marks JR, Hoey T, Lowe SW, Powers S: Oncogenic properties of ppm1d located within a breast cancer amplification epicenter at 17q23. Nat Genet 2002;31:133-134.
63 Lu X, Nguyen TA, Donehower LA. Reversal of the ATM/ATR- mediated DNA damage response by the oncogenic phosphatase PPM1D. Cell Cycle. 2005;4:1060–4.

64 Bulavin DV, Phillips C, Nannenga B, Timofeev O, Donehower LA, Anderson CW, Appella E, Fornace Jr AJ. Inactivation of the Wip1 phosphatase inhibits mammary tumorigenesis through p38 MAPK-mediated activation of the p16(Ink4a)-p19(Arf) pathway. Nat Genet. 2004;36:343–50.
65 Bulavin DV, Demidov ON, Saito S, Kauraniemi P, Phillips C, Amundson SA, Ambrosino C, Sauter G, Nebreda AR, Anderson CW, Kallioniemi A, Fornace AJ, Jr., Appella E: Amplification of ppm1d in human tumors abrogates p53 tumor-suppressor activity. Nat Genet 2002;31:210-215.
66 Demidov ON, Kek C, Shreeram S, Timofeev O, Fornace AJ, Appella E, Bulavin DV: The role of the mkk6/p38 mapk pathway in wip1-dependent regulation of erbb2-driven mammary gland tumorigenesis. Oncogene 2007;26:2502-2506.
67 Oliva-Trastoy M, Berthonaud V, Chevalier A, Ducrot C, Marsolier-Kergoat MC, Mann C, Leteurtre F: The wip1 phosphatase (ppm1d) antagonizes activation of the chk2 tumour suppressor kinase. Oncogene 2007;26:1449-1458.
68 Monz D, Munnia A, Comtesse N, Fischer U, Steudel WI, Feiden W, Glass B, Meese EU: Novel tankyrase-related gene detected with meningioma-specific sera. Clin Cancer Res 2001;7:113-119.
69 Wang S, Fusaro G, Padmanabhan J, Chellappan SP: Prohibitin co-localizes with rb in the nucleus and recruits n-cor and hdac1 for transcriptional repression. Oncogene 2002;21:8388-8396.
70 Rajalingam K, Rudel T: Ras-raf signaling needs prohibitin. Cell Cycle 2005;4:1503-1505.
71 Nijtmans LG, Artal SM, Grivell LA, Coates PJ: The mitochondrial phb complex: Roles in mitochondrial respiratory complex assembly, ageing and degenerative disease. Cell Mol Life Sci 2002;59:143-155.
72 Monte M, Collavin L, Lazarevic D, Utrera R, Dragani TA, Schneider C: Cloning, chromosome mapping and functional characterization of a human homologue of murine gtse-1 (b99) gene. Gene 2000;254:229-236.
73 Monte M, Benetti R, Buscemi G, Sandy P, Del Sal G, Schneider C: The cell cycle-regulated protein human gtse-1 controls DNA damage-induced apoptosis by affecting p53 function. J Biol Chem 2003;278:30356-30364.
74 Zhou X, Temam S, Oh M, Pungpravat N, Huang BL, Mao L, Wong DT: Global expression-based classification of lymph node metastasis and extracapsular spread of oral tongue squamous cell carcinoma. Neoplasia 2006;8:925-932.
75 Cekanova M, Fernando RI, Siriwardhana N, Sukhthankar M, De la Parra C, Woraratphoka J, Malone C, Strom A, Baek SJ, Wade PA, Saxton AM, Donnell RM, Pestell RG, Dharmawardhane S, Wimalasena J: Bcl-2 family protein, bad is down-regulated in breast cancer and inhibits cell invasion. Exp Cell Res 2015;331:1-10.

Legends to Figures

Fig 1 - A metastasis-associated gene expression signature. Fifty-eight genes and partial transcripts (ESTs) differentially expressed ($p \leq 0.01$) between invasive ductal breast tumor samples from patients with evidence of metastasis (n=9; full circles) and samples from patients that remain disease-free after at least 5-years follow-up (n=15; open circles) were identified following leave-one-out resampling (see *Materials* and *methods* for details). The genes (columns) are ordered by their correlation coefficient (signal-to-noise ratio) with the two patient outcome groups. Samples (rows) are ordered by their correlation to the metastasis profile, which is shown in the right panel. Expression level of each gene is represented by the number of standard deviations above (red) or below (green) the average value for that gene across all samples. Patients that remained metastasis-free have a correlation lower than 0.1 with the metastasis profile (dashed line in the right panel).

Fig 2 - Kaplan-Meier estimates of metastasis occurrence in patients with invasive ductal breast cancer based on the 58-gene metastasis signature(a), expression status of the Her2 receptor (b) or the Ki67 proliferation marker (c), and the histopathological grade (d).

Fig 3 - Kaplan-Meier estimates of metastasis occurrence in patients with invasive breast cancer based on public available data using the metastasis signature to generate risk scores (see Material and Methods for details). **a.** Two hundred ninety-four samples of primary breast carcinoma (data from [42]); **b**. Three hundred sixty-seven samples of breast carcinoma (data from [47]). The Log-rank test was used to ascertain the statistical difference between the survival curves.

Fig 4 – Systemic disease-free survival (SDFS, **a**) and cancer specific KM survival (CSS, **b**) curves according to BAD protein status in breast tumor samples. $p = 0.0001$ for SDFS and $p < 0.0001$ for CSS were determined by *log-rank* test.

Table 1 – Gene enrichment analysis using the 58-gene signature performed with Ingenuity Pathway Analysis (IPA) and KOBAS 2.0 tools.

Pathways	Database	ID	P-Value
RIP-mediated NFkB activation via ZBP1	IPA,Reactome	REACT_118563	0.002
TRAF6 mediated NF-kB activation	IPA,Reactome	REACT_24969	0.002
ZBP1(DAI) mediated induction of type I IFNs	IPA,Reactome	REACT_118764	0.003
TAK1 activates NFkB by phosphorylation of IKKs	IPA,Reactome	REACT_21281	0.003
p53 pathway	IPA,Reactome	REACT_121025	0.004
Nef Mediated CD8 Down-regulation	IPA,PANTHER	P00059	0.006
MyD88 cascade initiated on plasma membrane	IPA,Reactome	REACT_121175	0.009
Toll Like Receptor 5 (TLR5) Cascade	IPA,Reactome	REACT_118823	0.015
Toll Like Receptor 10 (TLR10) Cascade	IPA,KEGG	hsa05222	0.016
TRAF6 mediated induction of NFkB and MAP kinases	IPA,Reactome	REACT_25359	0.020
p53 pathway feedback loops 2	IPA,Reactome	REACT_11200	0.021
MyD88 dependent cascade initiated on endosome	IPA,Reactome	REACT_27215	0.022
Toll Like Receptor 7/8 (TLR7/8) Cascade	IPA,Reactome	REACT_9061	0.022
Toll Like Receptor 9 (TLR9) Cascade	IPA,Reactome	REACT_9027	0.022
Nef Mediated CD4 Down-regulation	IPA,Reactome	REACT_25024	0.023
EPH-Ephrin signaling	IPA,PANTHER	P04398	0.023
Toll Like Receptor TLR6:TLR2 Cascade	IPA,Reactome	REACT_25222	0.024
Toll Like Receptor 2 (TLR2) Cascade	IPA,Reactome	REACT_9020	0.024
Toll Like Receptor 3 (TLR3) Cascade	IPA,Reactome	REACT_9047	0.025
TRIF-mediated TLR3/TLR4 signaling	IPA,Reactome	REACT_11166	0.026
PECAM1 interactions	IPA,Reactome	REACT_6788	0.027
HDL-mediated lipid transport	IPA,Reactome	REACT_228170	0.027
p75NTR recruits signalling complexes	IPA,Reactome	REACT_8006	0.027
NF-kB is activated and signals survival	IPA,Reactome	REACT_8005	0.028
Activated TLR4 signalling	IPA,Reactome	REACT_7980	0.028
IRAK1 recruits IKK complex	IPA,Reactome	REACT_6783	0.030
WNT5A-dependent internalization of FZD4	IPA,Reactome	REACT_6809	0.030
Rho GTPase cycle	IPA,Reactome	REACT_25281	0.030
Signaling by Rho GTPases	IPA,Reactome	REACT_12519	0.031
p75NTR signals via NF-kB	IPA,Reactome	REACT_13621	0.034

Table 2 – The 58-gene metastasis signature was cross-referenced with lists of genes differentially expressed in metastatic breast cancer, retrieved from the data collected and processed at the NCBI-GEO database (http://www.ncbi.nlm.nih.gov/geo/) and originated from ten studies reported in the literature. Twelve genes that appeared in three or more studies were selected. Of these genes, 8 had the same signal direction for the fold change in more than 50% of studies in which expression values were measured. For each gene, expression values are shown and refer to log2 ratios between metastatic and non-metastatic breast tumors.

Gene	Description	This study	[a]TCGA	[b]van't Veer et al.	[c]van de Vijver et al.	[d]Desmedt et al.	[e]Minn et al.	[f]Zhao et al.	[g]Karlsson et al.	[h]Wang et al.	[i]Sørlie et al.	[j]Molly et al.	Jézéquel et al. [k]
GTSE1	G-2 and S-phase expressed 1	1.2	2.9	0.4	0.3	0.3				-0.1	-0.3	-0.7	
AP2B1	adaptor-related protein complex 2, beta 1 subunit	-1.4	0.06	-0.6			-0.2		-0.2	-0.2	-0.1	0.2	
ANGPTL1	angiopoietin-like 1	-1.5	-2.3		-0.5				0.4		0.3	0.5	
IKBKB	inhibitor of kappa light polypeptide gene enhancer in B-cells	-1.3	0.3				-0.3		0.9	-0.1	0.1	0.3	
FXYD1	FXYD domain-containing ion transport regulator 1	-1.0	-3.4			-0.2			-0.1	-0.1		0.1	0.4
SND1	Staphylococcal nuclease and tudor domain containing 1	0.4	0.07				0.2				-0.09	-0.1	0.1
ZNF664	zinc finger protein 664	0.5	0.07	0.6							0.1		
ITGAV	integrin, alpha V	1.0	-0.06					0.5	0.5		0.3	0.4	
NETO2	neuropilin (NRP) and tolloid (TLL)-like 2	1.2	0.8			0.3			-1.2	-0.8	-0.2	-0.5	
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha	0.5	0.3			-0.1		0.7					
EXDL2	exonuclease 3'-5' domain-like 2	0.4			-0.3					-0.06	0.1	0.1	
FKBP1B	FK506 binding protein 1B	1.6	-0.2			-0.1					-0.3	-0.4	

[a]TCGA (2012); [b]Nature (2002) 415:484; [c]N Engl J Med. (2002) 347:1999; [d]Clin Cancer Res. (2007) 13:3207; [e]Nature (2005) 436:518; [f]Mol Biol Cell. (2004) 15:2523; [g]BMC Cancer (2008) 8:254; [h]Breast Cancer Res (2011) 13(5):R92; [i]BMC Genomics (2006) 7:127; [j]PLoS One (2012) 7(2):e32426; [k]Breast Cancer Res Treat (2009) 116(3):509-20.

Table 3 – BAD protein expression in breast tumor samples detected by immunohistochemistry in tissue microarray and its association with clinical-pathological variables. Significant associations are in bold.

Variables	**N**	***BAD* expression score (%)**		***p*-value**
		negative	**positive**	
Age (years)				0.2440
<50	351	78 (22)	273 (78)	
≥50	557	106 (19)	451 (81)	
	908			
Familial history				0.3757
No	676	142 (21)	534 (79)	
Yes	183	33 (18)	150 (82)	
	859			
T				**0.0002**
T_1	76	13 (17)	63 (83)	
T_2	384	54 (14)	330 (86)	
T_3	128	33 (26)	95 (74)	
T_4	277	74 (27)	203 (73)	
	865			
N				**0.0035**
N_0	269	38 (14)	231 (86)	
N_1	265	54 (20)	211 (80)	
N_2	210	47 (22)	163 (78)	
N_3	141	41 (29)	100 (71)	
	885			
M				0.8366
Negative	816	167 (20)	649 (80)	
Positive	72	14 (19)	58 (81)	
	888			
Pathological Stage				**0.0002**
I	41	7 (17)	34 (83)	
II	359	49 (14)	310 (86)	
III	422	112 (27)	310 (74)	
IV	72	14 (19)	58 (81)	
Histological grade				0.6835
G1	120	21 (18)	99 (82)	
G2	532	112 (21)	420 (79)	
G3	259	53 (20)	206 (80)	
	911			
ER				**<0.0001**
Negative	279	86 (31)	193 (69)	
Positive	587	83 (14)	504 (86)	
	866			
PR				**<0.0001**
Negative	453	118 (26)	335 (74)	
Positive	384	43 (11)	341 (88)	
	837			
Her2				**0.0182**
Negative	654	147 (22)	507 (78)	
Positive	117	15 (13)	102 (87)	
	771			
Distant metastasis				**<0.0001**
No	432	65 (15)	367 (85)	
Yes	370	97 (26)	273 (74)	
	802			

T: Tumor; N: Nodes; M: Metastasis; ER: estrogen receptor alpha; PR: progesterone receptor; Her2: human epidermal growth factor receptor 2.