

# Prognostic Value of Gene Signatures and Proliferation in Lymph-Node-Negative Breast Cancer

Kristin Jonsdottir<sup>1</sup>, Jörg Assmus<sup>2</sup>, Aida Slewa<sup>1</sup>, Einar Gudlaugsson<sup>1</sup>, Ivar Skaland<sup>1</sup>, Jan P. A. Baak<sup>1,3</sup>, Emiel A. M. Janssen<sup>1\*</sup>

**1** Department of Pathology, Stavanger University Hospital, Stavanger, Norway, **2** Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway, **3** Free University, Amsterdam, The Netherlands

## Abstract

**Introduction:** The overall survival rate is good for lymph-node-negative breast cancer patients, but they still suffer from serious over- and some undertreatments. Prognostic and predictive gene signatures for node-negative breast cancer have a high number of genes related to proliferation. The prognostic value of gene sets from commercial gene-expression assays were compared with proliferation markers.

**Methods:** Illumina WG6 mRNA microarray analysis was used to examine 94 fresh-frozen tumour samples from node-negative breast cancer patients. The patients were divided into low- and high-risk groups for distant metastasis based on the MammaPrint-related genes, and into low-, intermediate- and high-risk groups based on the recurrence score algorithm with genes included in Oncotype DX. These data were then compared to proliferation status, as measured by the mitotic activity index, the expressions of phosphohistone H3 (PPH3), and Ki67.

**Results:** Kaplan-Meier survival analysis for distant-metastasis-free survival revealed that patients with weak and strong PPH3 expressions had 14-year survival rates of 87% ( $n = 45$ ), and 65% ( $n = 49$ ,  $p = 0.014$ ), respectively. Analysis of the MammaPrint classification resulted in 14-year survival rates of 80% ( $n = 45$ ) and 71% ( $n = 49$ ,  $p = 0.287$ ) for patients with low and high risks of recurrence, respectively. The Oncotype DX categorization yielded 14-year survival rates of 83% ( $n = 18$ ), 79% ( $n = 42$ ) and 68% ( $n = 34$ ) for those in the low-, intermediate- and high-risk groups, respectively ( $p = 0.52$ ). Supervised hierarchical cluster analysis for distant-metastasis-free survival in the subgroup of patients with strong PPH3 expression revealed that the genes involved in Notch signalling and cell adhesion were expressed at higher levels in those patients with distant metastasis.

**Conclusion:** This pilot study indicates that proliferation has greater prognostic value than the expressions of either MammaPrint- or Oncotype-DX-related genes. Furthermore, in the subgroup of patients with high proliferation, Notch signalling pathway genes appear to be expressed at higher levels in patients who develop distant metastasis.

**Citation:** Jonsdottir K, Assmus J, Slewa A, Gudlaugsson E, Skaland I, et al. (2014) Prognostic Value of Gene Signatures and Proliferation in Lymph-Node-Negative Breast Cancer. PLoS ONE 9(3): e90642. doi:10.1371/journal.pone.0090642

**Editor:** Paolo Provero, University of Turin, Italy

**Received:** December 20, 2012; **Accepted:** February 3, 2014; **Published:** March 5, 2014

**Copyright:** © 2014 Jonsdottir et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Kristin Jonsdottir has a grant from Helse Vest (no: 911450) (<http://www.helse-vest.no/english/Sider/side.aspx>), Norway and the study received financial support from the Folke Hermansen Foundation (2010) (<http://www.folke-fondet.org/index.html>) and the Stavanger University Hospital research department (2009–2010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Conflicts of interest: None of the authors have any financial or other relationships with entities that have investment, licensing, or other commercial interests in the subject matter under consideration in this article.

\* E-mail: emiel.janssen@sus.no

## Introduction

It is essential that the decision of whether or not to treat breast cancer patients is made as accurately as possible. It is therefore of utmost importance to be able to properly distinguish between those breast cancer patients who would benefit from adjuvant systemic therapy and those who could be spared such treatment. Different guidelines [1–3] have been developed to assist clinicians in making treatment decisions. These guidelines are often based on a combination of both clinical and pathological parameters that provide information about both the prognosis and prediction for therapy response. Unfortunately, even with these guidelines, over- and undertreatment of lymph node (LN)-negative breast cancer patients still occurs. For example, the St. Gallen guidelines advise adjuvant treatment in 85% of all node-negative breast cancers,

even though only 15–20% would die without treatment. Therefore, new parameters are needed to complement or replace the current clinicopathological features to ensure better characterization and treatment of breast tumours.

Microarray analysis can provide an expression profile of all genes in a tumour, thereby giving an impression of all of the active and inactive processes in that tumour. Studies on the gene-expression signatures of breast tumours have led to a new classification of breast cancers into at least five different subtypes with very different prognoses [4]. Supervised transcript profiling analyses have subsequently been used to develop standardized molecular prognostic indicators, such as MammaPrint and Oncotype DX [5,6]. These commercially available tests combine data from different biological pathways to provide information about both the prognosis and response to endocrine therapy and

chemotherapy. Oncotype DX low- and intermediate-risk patients have shown significant benefit from tamoxifen treatment, while patients with a high risk of distant breast cancer recurrence gain additional benefit from chemotherapy [7]. Interestingly, the genes that contributed most to these results were five proliferation genes (those encoding cyclin B1, Ki67, Myb-related protein B, survivin and serine/threonine-protein kinase 6), and those encoding the progesterone and oestrogen receptors (PR and ER, respectively) [6].

MammaPrint is a 70-gene expression assay that distinguishes between patients with high and low risks of distant recurrence using genes associated with proliferation, metastases, stromal invasion and angiogenesis. The first study was found to distinguish LN-negative breast cancer patients aged <55 years. The validations of this signature have included patients of all ages with both LN-positive and -negative disease, and it has been shown to predict relapse better than traditional histopathological features [8,9]. Interestingly, the gene signatures correlated with outcome contain high numbers of genes related to proliferation. A recent study found that a signature based upon cell-cycle-related genes alone was a more accurate predictor of breast cancer clinical outcome than another FDA-approved signature containing many more genes [10]. Another study [11] found that the simplest model for defining the risk score was the expression of a single proliferation gene, which yielded similar or an even better performance than models fitted from genome-wide data, and outperformed classical factors such as histological grade. These results are in agreement with previous retrospective and prospective studies involving large numbers of LN-negative breast cancer patients showing that proliferation measured by a thymidine labelling index, mitotic activity index (MAI), or phosphohistone H3 (PPH3) labelling are stronger prognosticators than classical predictors (reviewed in [12]). Moreover, two independent studies found that adjuvant chemotherapy was significantly beneficial for patients with rapidly proliferating tumours, but not for those with slowly proliferating tumours [13,14]. High rates of cell proliferation are correlated with shorter cancer survival [15,16]. Hence, it is of great interest to compare the prognostic value of the classical pathological prognosticators, proliferation markers, genes included in MammaPrint and Oncotype DX assays, and other new potential biomarkers for breast cancer.

The aim of this pilot study was, to determine whether gene expression can add prognostic information for subgroups of patients with tumours with low or high proliferative activity. Since proliferation measured by using MAI or PPH3 labelling has repeatedly proven to be the best prognosticator in LN-negative breast cancer (high sensitivity with little overtreatment), the rationale to commence with proliferation assessment for treatment decision-making may be logical.

## Material and Methods

### Patients

This study was approved by the Regional West committee for medical and health research ethics, the Norwegian Science Data Service and the Norwegian Data Inspectorate. None of the patients were required to provide written informed consent to participate since they had been diagnosed in the period 1993–1997. The regional ethics committee has approved this study. All insights in a patient's journal were monitored electronically, and all except the treating physician were required to state the reason why they needed to read that patient's journal. This log was always open for the patient to view. All patients were treated according to the national guidelines of the Norwegian Breast Cancer Group at

the time of diagnosis. Fresh-frozen tumour tissue for hormone receptor determination was collected from 235 individual patients, of which 135 patients were LN negative. Some of these patients had to be excluded because of bilateral disease ( $n=3$ ), previous breast tumours ( $n=8$ ), lack of follow-up ( $n=3$ ), lack of adequate material ( $n=10$ ) and poor-quality RNA ( $n=17$ ). This left 94 patients with adequate material and follow-up. Of these patients, 13 received endocrine therapy and 10 received chemotherapy. Of the 77 oestrogen receptor alpha (ER $\alpha$ )-positive patients, 12 received endocrine therapy and 4 received chemotherapy.

### Histopathology and Immunohistochemistry

The main tumour tissue was fixed in buffered 4% formaldehyde and then embedded in paraffin. Sections were cut at a thickness of 4  $\mu$ m and stained with haematoxylin, erythrosin and saffran. The histological type was assessed according to World Health Organization criteria [17], and the tumour grade was assessed according to the Nottingham modification [18]. The MAI and PPH3 labelling were assessed as described previously [19,20].

Immunohistochemistry (IHC) was used to detect ER $\alpha$ , progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), PPH3, Ki67 and cytokeratin 5/6 (CK5/6). The methods were based on DAKO technology as described previously [21]. ER $\alpha$  was scored positive if  $\geq 1\%$  of tumour cells exhibited nuclear staining, while all others were scored negative. PR was scored as positive when nuclear staining was present in  $\geq 10\%$  and scored negative when  $< 10\%$  of the tumour cells had nuclear staining. The DAKO Hercep-Test scoring protocol was used for measuring HER2, with cases scored as 3+ considered to be positive. The cases that were scored as 2+ were further validated with fluorescence in situ hybridization using the Pathvysion HER2 DNA probe kit (Abbott Laboratories, Abbott Park, IL, USA) for HER2-neu amplification. The manufacturer's protocol was followed. Triple-negative breast cancers (TNPs) were defined as being negative for ER $\alpha$  (0%), PR ( $< 10\%$ ) and HER2 (– and +). The semiautomatic interactive computerized QPRODIT system (Leica, Cambridge, UK) was used for measuring the percentage of Ki67-positive cells, as described by Gudlaugsson et al. [22]. CK5/6-positive tumour cells were scored using a continuous scale of 0–100%, where in the final analysis all tumours with any CK5/6 staining in tumour cells were grouped as being positive, as described previously [23]. All sections were independently scored by two pathologists.

### RNA Isolation/Labeling/Hybridization

All cryosections used for RNA isolations were evaluated by an experienced breast pathologist (E.G.). An area comprising at least 50% tumour cells was isolated by means of macrodissection. At least two 10- $\mu$ m cryosections were used for total RNA isolation using the MirVANA total RNA isolation kit (Ambion/Applied Biosystems, Austin, TX, USA), according to the protocol provided by the manufacturers. For quality control, all samples were analysed using both the Agilent 2100 Bioanalyzer system (total RNA and small RNA chips) and the NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The microarray experiment was performed using the Illumina iScan device, which uses fluorescence detection of biotin-labelled cRNA. For each sample, 250 ng of total RNA was reverse transcribed, amplified and labelled with biotin-UTP using an Illumina TotalPrep-96 RNA Amplification Kit (version 4393543, Ambion/Applied Biosystems). The quantity of labelled cRNA was measured using the NanoDrop spectrophotometer (Thermo Scientific), whereas the quality and size distribution of the labelled cRNA was assessed using the 2100 Bioanalyzer (Agilent). Finally, 1.5  $\mu$ g of

biotin-labelled cRNA was hybridized to Illumina HumanWG-6 v3 Expression BeadChips according to the manufacturer's protocol.

### Survival Endpoints

Distant-metastasis-free survival (DMFS) was used as the main endpoint for both Kaplan-Meier survival plots and as an event for both univariate and multivariate Cox analysis. Patients were surveyed until the date of last follow-up visit for death from causes other than breast cancer, local or regional recurrences, and the development of a second primary cancer, including contralateral breast cancer. If a patient's status during follow-up indicated a confirmed metastasis without a recurrence date, the last follow-up visit date was used. Age, time to first distant recurrence and survival time were calculated relative to the primary diagnosis date.

### Statistical Analyses

Statistical analyses were conducted using SPSS (version 20.0, SPSS, Chicago, IL, USA) and MATLAB 7.10 (MathWorks, Natick, MA, USA). Differences between patient groups were tested using the log-rank test and Kaplan-Meier survival curves. Cox proportional-hazard analysis (forward, Wald) and hazard ratio (HR) with 95% confidence intervals (CIs) were used to determine the relative importance of the prognostic variables.

The MammaPrint classification and recurrence score (RS) algorithm from Oncotype DX were computed as described by Paik et al. [6,8] and van de Vijver et al. [6,8]. The average gene-expression profile of patients (AGPP) with and without distant metastasis was calculated using MammaPrint-related genes. Each sample was then correlated with each of these AGPPs and classified. Receiver operating characteristic (ROC) curve analysis was used with respect to DMFS to determine cut-offs for both indices. The optimal cut-offs were defined as those with an ROC analysis value closest to 1. For Oncotype DX, a second cut-off was defined as the optimal cut-off in the highly sensitive part of the curve, thus defining three risk categories: low, medium and high. In addition, a cluster analysis was performed for the tumours using both Pearson and Spearman correlations, and the two main clusters were defined as high- and low-risk groups. Kaplan-Meier survival analysis was used with the log-rank test to test whether there was a significant difference in DMFS between the different patient groups determined univariately by different risk factors (Table 1). A multivariate analysis was performed including all significant univariate factors from Table 1, according to a Cox regression. This data set is available publicly in the Gene Expression Omnibus: accession number GSE46563.

The free downloadable software package dChip (version 31 March 2009) was used to identify genes related to different clinical features and to the presence or not of distant metastases (survival time was not taken into account). Absolute correlations (including genes with opposing gene profiles) were calculated using ANOVA, with the cut-off for statistical significance set at  $p \leq 0.01$ . Gene lists created in this way were used to classify samples by cluster analysis. The classification accuracy was tested by performing a cross-validation analysis (by leave one sample out and reselect genes). The created gene lists were further analysed for Gene Ontology terms using the free software programs Gorilla [24,25] and DAVID [26,27]. The same exercise was also performed for the patients with high proliferation only.

### Results

Ninety-four patients with LN-negative breast cancer were included; their clinical features are listed in Table 1. The median

**Table 1.** Distant-metastasis-free survival in lymph node negative breast cancer patient with Kaplan-Meier survival- and univariate analysis.

Characteristic	Distant metastases			
	Event/at risk (%)	Log-rank P-value	HR	95% CI
<b>Age</b>				
<55 years	6/35 (83)	0.099	2.2	0.8–5.5
≥55 years	17/59 (71)			
<b>Tumour diameter</b>				
<2 cm	12/56 (79)	0.334	1.9	0.8–4.3
≥2 cm	11/38 (71)			
<b>Nottingham grade</b>				
1	3/14 (79)	0.759	1.2	0.4–4.1
2 and 3	20/80 (75)			
<b>ER</b>				
Positive ≥1%	18/77 (77)	0.585	1.3	0.5–3.6
Negative	5/17 (71)			
<b>PR</b>				
Positive ≥10%	17/78 (78)	0.268	1.4	0.6–3.2
Negative 0–9%	6/16 (63)			
<b>HER2</b>				
0–1+	21/82 (74)	0.533	1.6	0.4–6.7
2+–3+	4/14 (83)			
<b>MAI</b>				
<10	11/60 (82)	0.041	2.3	1.0–5.2
≥10	12/34 (65)			
<b>MAI</b>				
MAI 0–2	2/34 (94)	0.003	6.7	1.5–31.8
MAI 3–9	9/26 (65)			
MAI ≥10	12/34 (65)			
<b>Ki67</b>				
0–9%	5/39 (87)	0.05	2.6	1.0–7.0
10–100%	18/53 (66)			
<b>PPH3</b>				
<13	6/45 (87)	0.014	3.1	1.2–7.8
≥13	17/49 (65)			
<b>CK5/6</b>				
Negative	19/82 (77)	0.547	1.4	0.5–6.0
Positive	4/12 (67)			
<b>Triple Negative</b>				
Positive	21/89 (76)	0.352	1.4	0.5–4.2
Negative	2/5 (60)			
<b>Oncotype DX genes</b>				
Low risk	3/18 (83)	0.522	1.2	0.3–4.4
Medium risk	9/42 (79)			
High risk	11/34 (68)			
<b>MammaPrint genes</b>				
Good prognosis	9/45 (80)	0.287	1.6	0.7–3.6
Bad prognosis	14/49 (71)			

doi:10.1371/journal.pone.0090642.t001

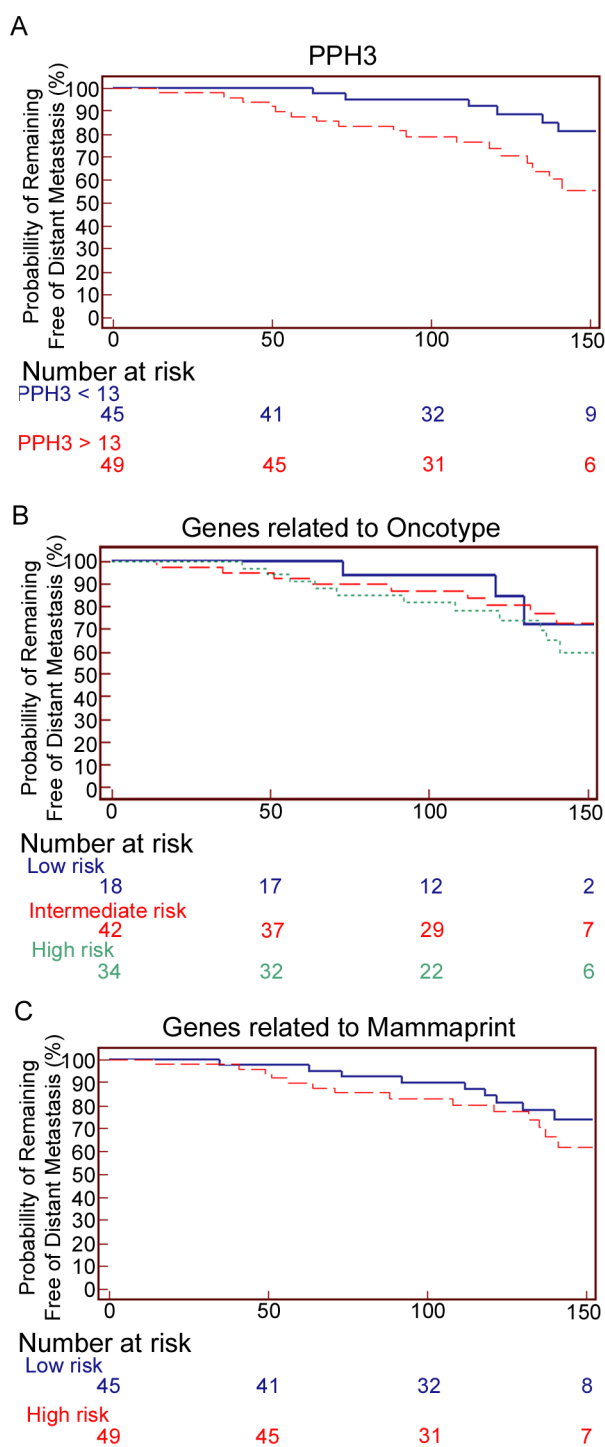
age of this group was 60 years, and the median follow-up period was 127 months (range 14–171 months). Twenty-three patients (24%) developed distant metastasis or died from breast-cancer-related disease. The genes involved in the Oncotype DX [6] and the MammaPrint assay [28], and their concordant Illumina names, are presented in Tables S1 and S2 in File S1. Of the 70 genes present in the MammaPrint assay, 62 were identified in the Illumina array, as were all of the genes included in the Oncotype DX assay. Table 1 presents the DMFS and HRs for all of the tumour characteristics; MAI, and PPH3 and Ki67 status appear to be strong prognostic factors in this cohort of samples.

Kaplan-Meier survival analysis revealed that a high expression of PPH3 significantly identified patients who develop distant metastasis (Figure 1). Patients with high and low expressions of PPH3 had 14-year survival rate of a 65% and 87%, respectively ( $p=0.014$ , HR = 3.1, 95% CI = 1.2–7.8). To test the prognostic value of the MammaPrint genes, the patients were classified into groups with low and high risks of DMFS based on the expression of these genes using hierarchical cluster analysis (Figure S1). Kaplan-Meier survival analysis of MammaPrint grouping resulted in a DMFS rates of 80% and 71% for low- and high-risk patients, respectively ( $p=0.287$ , HR = 1.6, 95% CI = 0.7–3.6) with a 14-year follow-up. Furthermore, to validate the prognostic value of the 16 target genes related to Oncotype DX classification, the patients were divided into low-, intermediate- and high-risk groups according to their expression of these genes and based on ROC-curve analysis of RS (Figure S2); this categorization resulted in survival rates of 83% ( $n=18$ ), 79% ( $n=42$ ) and 68% ( $n=34$ ), respectively ( $p=0.522$ , HR = 1.2 and 1.8, 95% CI = 0.3–4.4 and 0.5–6.5) with a 14-year follow-up. Originally, the Oncotype DX assay was designed for use in ER $\alpha$ -positive patients. Therefore Kaplan-Meier survival analysis was also performed in the subgroup of ER $\alpha$ -positive patients ( $n=77$ ), which yielded survival rates of 83%, 81% and 61% in the low-, intermediate- and high-risk groups, respectively ( $p=0.293$ ).

ANOVA-analysis of all genes included in the Illumina WG6 array versus presence of distant metastasis or not (survival time was not taken into account) revealed that 82 genes were significantly ( $p=0.01$ ) associated with the presence or not of distant metastases (Figure 2); these genes are listed in Table S3 in File S1. The cross-validation (by leaving one sample out and reselecting genes) of overall classification accuracy was found to be 67% for these 82 genes. The same analysis showed that the sensitivity and specificity were 70% and 57%, respectively. The known biological processes associated with these 82 genes are also listed in Table 2. Genes related to pregnancy, fatty-acid metabolic processes and the regulation of growth are included in this signature. Furthermore, three of the genes were also found among the genes from MammaPrint (those encoding endothelial cell-specific molecule 1, origin recognition complex subunit 6 and Ras-related protein Rab-6B: ESM1, ORC6L and RAB6B, respectively).

In the same way, investigation of which genes were significantly associated with the presence or not of distant metastases in patients with high proliferation (PPH3 $\geq$ 13) revealed that genes involved in Notch signalling and cell adhesion were expressed at higher levels in patients with distant metastases (Table 3). Seventeen genes from this gene list were also included in the 82-gene signature list, while none were found among the MammaPrint or Oncotype DX genes. No significant pathways were identified in the DMFS-related genes in the 45 patients with low proliferation, among whom only 6 patients developed distant metastasis.

Both forward and backward multivariate analysis of all significant features from Table 1 showed that PPH3 was the strongest independent prognostic factor.

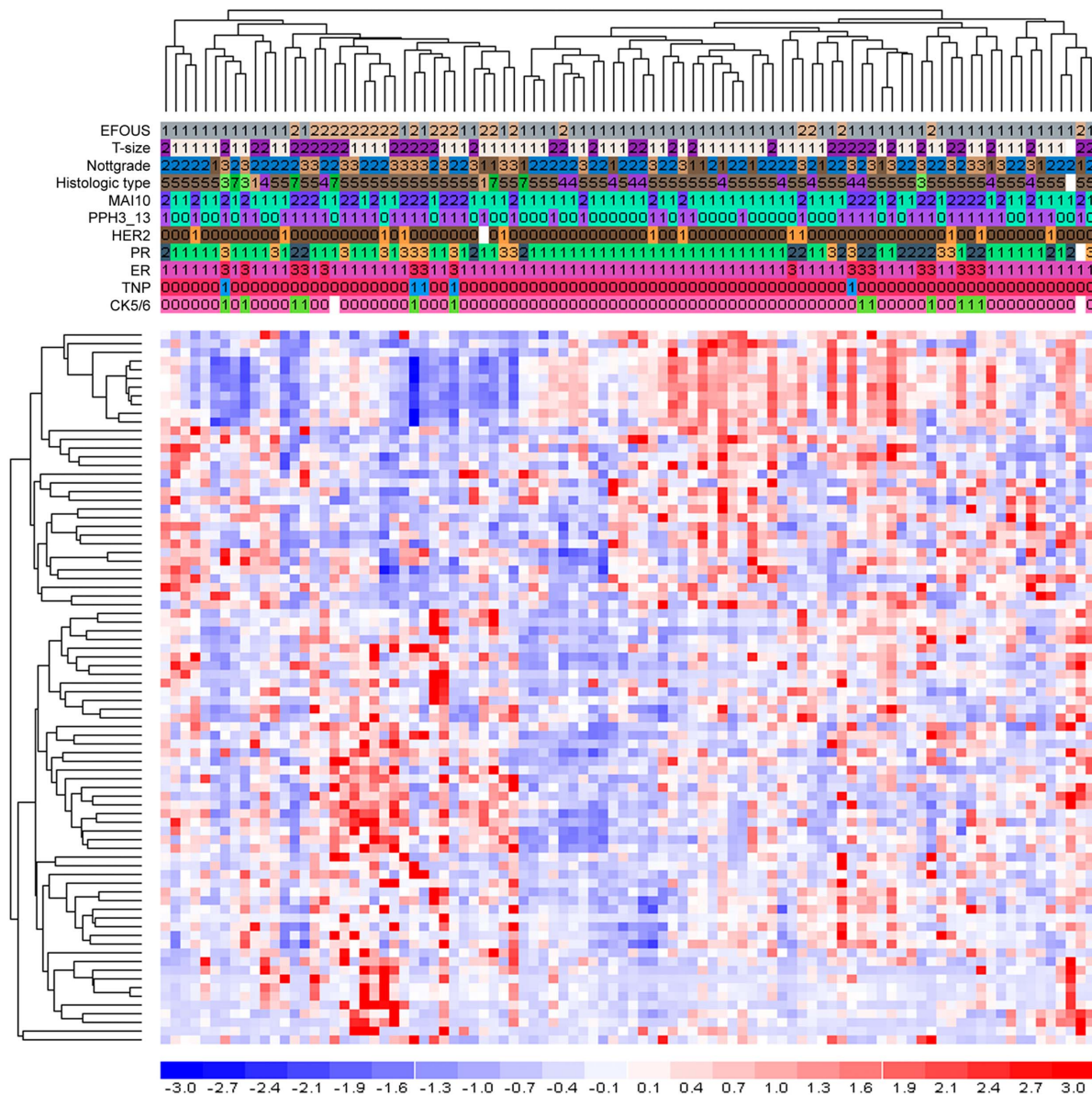


**Figure 1. Long-term distant-metastasis-free survival curves according to PPH3 expression (A), classification by Oncotype DX RS (B), and classification by genes related to MammaPrint (C).**

doi:10.1371/journal.pone.0090642.g001

## Discussion

This pilot study compared gene-expression data with the proliferation markers MAI, Ki67 and PPH3 status. Proliferation was the strongest independent prognostic factor in this cohort of breast cancer patients. Furthermore, in the group of patients with



**Figure 2. Supervised hierarchical clustering for DMFS.** Each row represents an mRNA and each column represents a patient sample. The mRNA clustering tree is shown on the left, and the sample clustering tree appears at the top. The colour scale shown at the bottom illustrates the relative expression level of an mRNA across all samples: red colour represents an expression level above mean, blue colour represents expression lower than the mean. Gray colour means that the specific mRNA has not been successfully detected with microarray. Numbers for clinicopathological features indicate the following: EOFUS (1 = no distant metastasis, 2 = distant metastasis), Tsize (Tumour size: 1  $\leq$  2 cm, 2 > 2 cm) and Nottgrade (Nottingham grade: 1 = grade 1, 2 = grade 2, 3 = grade 3), Histologic type (1 = Tubular, 2 = Colloid, 3 = Medullary, 4 = Lobular, 5 = Ductal, 6 = mix Ductal/Lobular, 7 = Others), MAI10 (1 < 10, 2  $\geq$  10), PPH3\_13 (0 < 13, 1  $\geq$  13), HER2 (0 = 0 or 1+, 1 = 2+ or 3+), PR (1 < 1%, 2 = 1–9% positive tumour cells, 3  $\geq$  10% positive tumour cells), ER $\alpha$  (1 < 1% positive tumour cells, 3  $\geq$  1% positive tumour cells), TNP (0 = positive for either ER $\alpha$ /PR/Her2, 1 = negative for ER $\alpha$  and PR and HER2), and CK5/6 (0 = no staining, 1 = any percentage of positive tumour cells). doi:10.1371/journal.pone.0090642.g002

highly proliferative tumours, several genes related to the Notch signalling pathway appeared to add prognostic value.

Investigating expression data of groups of genes rather than single genes is believed to be a more reliable indicator of clinical response. Retrospective studies of the Oncotype DX assay have shown that it predicts recurrence better than classical clinicopathological variables [7,29]. The assay is therefore widely requested by physicians, especially in North America, and studies have shown that the application of this assay changes patient management in 30% of cases [30]. The assay is currently only

available in centralized laboratories, and is much more expensive than IHC. Evaluation of the Oncotype DX assay with a combined ER, PR, HER2 and Ki67 IHC score revealed a correlation coefficient of 0.7 [31]. Direct comparison of mRNA level [measured with quantitative polymerase chain reaction (qPCR)] and protein levels of ER, PR and HER2 (measured with IHC) revealed a high concordance value (79–94%) [32], while several other studies have found that a proliferation score based on PPH3 or Ki67 counts was significantly correlated with RS [33–35]. On the other hand, some have suggested that none of the standard

**Table 2.** List of gene-ontology terms related to 82-gene signatures and presence or not of distant metastases.

Term	Genes involved	P-Value
Female pregnancy	PGF; placental growth factor;PSG3; pregnancy specific beta-1-glycoprotein 3; PSG4; pregnancy specific beta-1-glycoprotein 7; pregnancy specific beta-1-glycoprotein 8; pregnancy specific beta-1-glycoprotein 4 RG Homo sapiens; PSG9 pregnancy specific beta-1-glycoprotein 9; TRO; trophinin	7.1E-4
Fatty acid metabolic process	ACSL5; acyl-CoA synthetase long-chain family member 5; ADIPOR1; adiponectin receptor 1; C9ORF3; chromosome 9 open reading frame 3; PLA2G10; phospholipase A2, group X	3.7E-2
Regulation of growth	NELL2; NEL-like 2 (chicken); ADIPOR1; adiponectin receptor 1; ESM1; endothelial cell-specific molecule 1; FOXS1; forkhead box S1 RG Homo sapiens; TRO; trophinin	3.7E-2
rRNA processing	EXOSC7; exosome component 7; RPL11 ribosomal protein L11; RPS7; ribosomal protein S7; ribosomal protein S7 pseudogene 11; ribosomal protein S7 pseudogene 4; ribosomal protein S7 pseudogene 10	4.5E-2
rRNA metabolic process	EXOSC7; exosome component 7 RG; RPL11; ribosomal protein L11; RPS7; ribosomal protein S7; ribosomal protein S7 pseudogene 11; ribosomal protein S7 pseudogene 4; ribosomal protein S7 pseudogene 10	4.9E-2
Ribosome biogenesis	EXOSC7; exosome component 7; RPL11; ribosomal protein L11; RPS7; ribosomal protein S7; ribosomal protein S7 pseudogene 11; ribosomal protein S7 pseudogene 4; ribosomal protein S7 pseudogene 10	7.5E-2
Defence response	F12; coagulation factor XII (Hageman factor); CCR6; cyclin L2; chemokine (C-C motif) receptor 6; LILRB3; leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3; PSG3; pregnancy specific beta-1-glycoprotein 3; PSG4; pregnancy specific beta-1-glycoprotein 7; pregnancy specific beta-1-glycoprotein 8; pregnancy specific beta-1-glycoprotein 4; PSG9; pregnancy specific beta-1-glycoprotein 9	7.6E-2
Cellular response to hormone stimulus	ADIPOR1 adiponectin receptor 1; GCGR; glucagon receptor; PGF; placental growth factor	8.7E-2

doi:10.1371/journal.pone.0090642.t002

clinicopathological features can accurately predict the RS [36]. Nevertheless, a meta-analysis of 647 ER $\alpha$ -positive patients with information about clinicopathological features and the RS demonstrated that the combined use of all of this information was more powerful than RS alone [37]. This combination also resulted in fewer patients being categorized in the intermediate-risk group, whereas up to 66% were classified as being at intermediate risk by RS alone. The present analysis of Oncotype-

DX-related genes showed that although the survival rate appeared to be lower and the number of patients who suffered a recurrence appeared to be higher for the high-risk group than for the low-risk group, the differences were not statistically significant.

There are some technical differences between the present study and the original Oncotype DX assay, which was designed to detect the gene expression of mRNA isolated from formalin-fixed, paraffin-embedded (FFPE) material by means of qPCR. In the

**Table 3.** List of gene-ontology terms related high expression of PPH3 and presence or not of distant metastases.

Term	Genes involved	P-Value
Notch signalling pathway	NOTCH2NL; Notch homolog 2 N-terminal like; NOTCH4; Notch homolog 4; HEYL; hairy/enhancer-of-split related with YRPW motif-like	9.8E-3
Regulation of muscle contraction	ARG2; arginase, type II RG Homo sapiens; MYL5; myosin, light chain 5, regulatory; TNNT1; troponin T type 1 (skeletal, slow)	1.8E-2
Cell morphogenesis involved in differentiation	NOTCH4; Notch homolog 4; NDN; necdin homolog; NTN1; netrin 1; PLA2G10; phospholipase A2, group X	3.3E-2
Negative regulation of muscle contraction	ARG2; arginase, type I; TNNT1; troponin T type 1 (skeletal, slow)	3.4E-2
Cell-cell adhesion	CD93; CD93 molecule; FAT3; FAT tumour suppressor homolog 3; ESAM; endothelial cell adhesion molecule; PCDHB12; protocadherin beta 12	4.5E-2
Cell adhesion	CD93; CD93 molecule; FAT3; FAT tumour suppressor homolog 3; NELL2; NEL-like 2; COL5A3; collagen, type V, alpha 3 ESAM; endothelial cell adhesion molecule; PCDHB12 protocadherin beta 12	4.9E-2
Biological adhesion	CD93; CD93 molecule; FAT3; FAT tumour suppressor homolog 3; NELL2; NEL-like 2; COL5A3; collagen, type V, alpha 3; ESAM; endothelial cell adhesion molecule; PCDHB12; protocadherin beta 12	4.9E-2
Homophilic cell adhesion	FAT3; FAT tumour suppressor homolog 3; ESAM endothelial cell adhesion molecule; PCDHB12 protocadherin beta 12	5.5E-2
Regulation of system process	S100P S100 calcium binding protein P; ARG2 arginase, type II; MYL5 myosin, light chain 5, regulatory; TNNT1 troponin T type 1 (skeletal, slow)	5.9E-2
Regulation of growth	NELL2; NEL-like 2; FOXS1; forkhead box S1; NDN; necdin homolog; NTN1; netrin 1	7.5E-2
Cell morphogenesis	NOTCH4; Notch homolog 4; NDN; necdin homolog; NTN1; netrin; PLA2G10; phospholipase A2, group X	8.2E-2
Cell rejection organization	FGD5; FYVE, RhoGEF and PH domain containing 5; NDN; necdin homolog; NTN1; netrin 1; PLA2G10; phospholipase A2, group X	8.9E-2

doi:10.1371/journal.pone.0090642.t003

current study mRNA was isolated from fresh-frozen tumours and the gene expression was analysed using microarray technology. These technical differences do not seem to impair comparisons between the survival rates, since a paired comparison of gene expression between FFPE and fresh-frozen tissue yielded a strong correlation [38]. Furthermore, a strong correlation was also observed between microarray and qPCR findings [39], indicating that the present results are comparable to those of other studies that have employed Oncotype DX.

The 70-gene MammaPrint signature can divide patients into two separate groups: those with low and high risks of distant metastasis. The first study of this signature showed very good separation of the patients into these two groups, with HR and 95% CI values for the time to distant metastasis of 4.6 and 2.3–9.2, respectively, at a 10-year follow-up [8]. A validation study of the 70-gene profile signature showed that the assay is strongly prognostic, although the magnitude of the effect was much less than previously reported, with HR and 95% CI values for the time to distant metastases now being 2.32 and 1.35–4.00, respectively [40]. By comparison, a large multicentre prospective study exploring the prognostic value of the MAI found HR and 95% CI values for remaining free of distant metastasis of 3.12 and 2.17–4.50, respectively, thereby exceeding the prognostic value of the 70-gene signature [41]. Analyses of the prognostic information that lies in the 70-gene signature and other multigene signatures have shown that a large portion of the prognostic information lies in proliferation-related genes [42]. In fact, reanalyses of these signatures showed that the signature with proliferation-related genes had greater prognostic value than the original signature [43–45], and that the proliferation signature was correlated with the MAI (correlation coefficient, 0.968) [44]. One study showed that the non-proliferative genes had no prognostic power [43]. Reanalysis of the data from van't Veer et al. [5] showed that the molecular signature was strongly dependent on the selection of patients and that the set of genes was therefore not unique [46]. The present results with MammaPrint-related genes suggest that the proportion of patients with recurrence would be higher in the high-risk group than in the low-risk group. However, none of the analyses produced statistically significant results. The original MammaPrint assay was performed at a centralized laboratory and with a different microarray platform. Differences in the platforms used could influence the results; however, one study has shown that the Illumina and Affimetrix platforms are highly comparable [47]. Differences in the computation of the risk scores and definitions of thresholds by van't Veer et al. and in the present study could also be a possible reason for some of the discrepancies between the findings. Furthermore, not all of the genes were identified in the MammaPrint assay, which could have affected the statistical analysis.

Both the MammaPrint and Oncotype DX assays are based on RNA isolation from the tumour area, so that the harvested RNA is likely to be a mixture of both cancerous and normal tissue, and/or other non-cancerous cells. This may have affected the gene-expression profile and introduced significant bias into the prediction of the patients' outcome [48–50]. Furthermore, the particular patient cohort included can affect the analysis and result in different gene signatures, as was the case in the current study; the gene signature of this study contained only three genes that were included in the MammaPrint signature, thereby confirming the observation that many different signatures can provide similar prognostic information, even in the same patient cohort [51].

Fresh-frozen tumour material for research is generally harvested from palpable and often larger tumours. Since tumour size is known to be related to prognosis, this may explain why the present

cohort included a higher percentage of patients who developed distant metastasis.

The proliferation marker PPH3 is known to be a strong prognostic marker for LN-negative breast cancer; 35% of 49 patients with a high PPH3 suffered a recurrence. Although this means a high risk of developing distant metastasis, it also means that 65% did not develop distant metastasis. Therefore, it would be of great interest to investigate whether gene-expression analyses could add prognostic value to markers of proliferation. This could help to refine which patients really need chemotherapy, thereby decreasing both under- and overtreatment. In the present study, genes related to the Notch signalling pathway and Notch genes were expressed at significantly higher levels in patients with high proliferation rates who developed distant metastasis. The Notch family of cell-surface receptors is involved in proliferation, migration and invasion [52,53]. Notch receptors have also been shown to regulate the self-renewal of mammary stem cells [54]. Furthermore, high Notch expression is observed in TNP, and is correlated with an overall poor outcome [55,56]. In vivo knockdown experiments of Notch 1 and Notch 4 in tumour-bearing mice showed that the tumour size decreased and there was a reduction of breast cancer recurrence in these mice [53,57]. Cell-line and in vivo experiments with ER $\alpha$ -positive breast cancer cells showed that tamoxifen treatment reactivated Notch signalling, which again induced proliferation and invasion. The same study showed that a combination of inhibition of Notch signalling and tamoxifen treatment inhibits tumour growth [52]. This observation has been confirmed by others [58].

In conclusion, the findings of this pilot study indicate that markers of proliferation status outperform MammaPrint- and Oncotype-DX-related genes as prognostic markers for LN-negative breast cancer. The inclusion of separate measurements of proliferation in future microarray studies might therefore be warranted. Furthermore, the obtained data support the previous finding that Notch could be a potential prognostic and predictive marker for the subgroup of highly proliferative LN-negative breast cancer patients.

## Supporting Information

**Figure S1** Unsupervised cluster analysis with the genes related to MammaPrint assay. Colour codes: In the heat map (green colour indicate low expression of the mRNA and red indicates high expression), cluster prognosis (red = high risk of distant-metastasis, green = low risk of distant-metastasis), Distant metastasis (red = developing distant metastasis, green = no distant-metastasis), and correlation to average gene expression profile (AGPP) for no distant metastasis. (TIF)

**Figure S2** ROC-curve analysis for Oncotype DX RS and the optimal cut-offs. (TIF)

**File S1** Table S1, List and description of genes related to Oncotype DX. Table S2, List and description of genes related to MammaPrint assay. Table S3, List of 82-genes significant associated to DMFS. (DOC)

## Acknowledgments

The microarray experiments were performed at the Norwegian Microarray Consortium at the University of Oslo, Oslo, which is supported by the FUGE programme of The Norwegian Research Council.

## Author Contributions

Conceived and designed the experiments: KJ EAMJ JPAB. Performed the experiments: AS. Analyzed the data: JA KJ EAMJ. Wrote the paper: KJ

## References

- Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, et al. (2011) Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 22: 1736–1747.
- The Norwegian Breast Cancer Group. Available: <http://www.nbcg.no/index.html>. Accessed 2012 Dec. 10.
- Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, et al. (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25: 5287–5312.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. *Nature* 406: 747–752.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530–536.
- Paik S, Shak S, Tang G, Kim C, Baker J, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351: 2817–2826.
- Paik S, Tang G, Shak S, Kim C, Baker J, et al. (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 24: 3726–3734.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999–2009.
- Knauer M, Mook S, Rutgers EJ, Bender RA, Hauptmann M, et al. (2010) The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat* 120: 655–661.
- Git A, Spiteri I, Blenkinsop C, Dunning MJ, Pole JC, et al. (2008) *PMCA2*, a breast progenitor cancer cell line, has normal-like mRNA and microRNA transcriptomes. *Breast Cancer Res* 10: R54.
- Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, et al. (2008) Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 14: 5158–5165.
- Baak JP, Gudlaugsson E, Skaland I, Guo LH, Klos J, et al. (2009) Proliferation is the strongest prognosticator in node-negative breast cancer: significance, error sources, alternatives and comparison with molecular prognostic markers. *Breast Cancer Res Treat* 115: 241–254.
- Baak JP, Kruse AJ, Robboy SJ, Janssen EA, van Diermen B, et al. (2006) Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. *Journal of Clinical Pathology* 59: 1017–1028.
- Andre F, Khalil A, Slimane K, Massard C, Mathieu MC, et al. (2005) Mitotic index and benefit of adjuvant anthracycline-based chemotherapy in patients with early breast cancer. *Journal of Clinical Oncology* 23: 2996–3000.
- Baak JP, van Diest PJ, Voorhorst FJ, van der Wall E, Beex LV, et al. (2007) The prognostic value of proliferation in lymph-node-negative breast cancer patients is age dependent. *Eur J Cancer* 43: 527–535.
- Nicholson RI, Bouzbar N, Walker KJ, McClelland R, Dixon AR, et al. (1991) Hormone sensitivity in breast cancer: influence of heterogeneity of oestrogen receptor expression and cell proliferation. *Eur J Cancer* 27: 908–913.
- Tavassoli FA, Devilee P, (Eds.) (2003) World Health Organization Classification of Tumors Pathology and Genetics of Tumors of the Breast and Female and Genital Organs: oIARC Press: Lyon.
- Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19: 403–410.
- Baak JP, van Diest PJ, Voorhorst FJ, van der Wall E, Beex LV, et al. (2005) Prospective multicenter validation of the independent prognostic value of the mitotic activity index in lymph node-negative breast cancer patients younger than 55 years. *Journal of Clinical Oncology* 23: 5993–6001.
- Skaland I, Janssen EA, Gudlaugsson E, Klos J, Kjelleveid KH, et al. (2007) Phosphohistone H3 expression has much stronger prognostic value than classical prognosticators in invasive lymph node-negative breast cancer patients less than 55 years of age. *Modern Pathology* 20: 1307–1315.
- Skaland I, Janssen EA, Gudlaugsson E, Klos J, Kjelleveid KH, et al. (2009) Validating the prognostic value of proliferation measured by Phosphohistone H3 (PPH3) in invasive lymph node-negative breast cancer patients less than 71 years of age. *Breast Cancer Res Treat* 114: 39–45.
- Gudlaugsson E, Skaland I, Janssen EA, Smaaland R, Shao Z, et al. (2012) Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. *Histopathology*.
- Janssen EA, Ovestad IT, Skaland I, Soiland H, Gudlaugsson E, et al. (2009) LOH at 1p31 (ARHI) and proliferation in lymph node-negative breast cancer. *Cell Oncol* 31: 335–343.
- Eden E, Lipson D, Yogeve S, Yakhini Z (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol* 3: e39.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48.
- Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57.
- Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13.
- Tian S, Roepman P, Van't Veer LJ, Bernards R, de Snoo F, et al. (2010) Biological functions of the genes in the mammaprint breast cancer profile reflect the hallmarks of cancer. *Biomark Insights* 5: 129–138.
- Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, et al. (2006) A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 8: R25.
- Lo SS, Mumby PB, Norton J, Rychlik K, Smerage J, et al. (2010) Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. *J Clin Oncol* 28: 1671–1676.
- Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, et al. (2011) Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol* 29: 4273–4278.
- Sun B, Zhang F, Wu SK, Guo X, Zhang LL, et al. (2011) Gene expression profiling for breast cancer prognosis in Chinese populations. *Breast J* 17: 172–179.
- Sahebjam S, Aloyz R, Pilavdzic D, Brisson ML, Ferrario C, et al. (2011) Ki 67 is a major, but not the sole determinant of Oncotype Dx recurrence score. *Br J Cancer* 105: 1342–1345.
- Niikura N, Iwamoto T, Masuda S, Kumaki N, Xiaoyan T, et al. (2012) Immunohistochemical Ki67 labeling index has similar proliferation predictive power to various gene signatures in breast cancer. *Cancer Sci* 103: 1508–1512.
- Williams DJ, Cohen C, Darrow M, Page AJ, Chastain B, et al. (2011) Proliferation (Ki-67 and phosphohistone H3) and oncotype DX recurrence score in estrogen receptor-positive breast cancer. *Appl Immunohistochem Mol Morphol* 19: 431–436.
- Wolf I, Ben-Baruch N, Shapira-Frommer R, Rizel S, Goldberg H, et al. (2008) Association between standard clinical and pathologic characteristics and the 21-gene recurrence score in breast cancer patients: a population-based study. *Cancer* 112: 731–736.
- Kelly CM, Krishnamurthy S, Bianchini G, Litton JK, Gonzalez-Angulo AM, et al. (2010) Utility of oncotype DX risk estimates in clinically intermediate risk hormone receptor-positive, HER2-normal, grade II, lymph node-negative breast cancers. *Cancer* 116: 5161–5167.
- Sanchez-Navarro I, Gamez-Pozo A, Gonzalez-Baron M, Pinto-Marin A, Hardisson D, et al. (2010) Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen and formalin-fixed, paraffin-embedded breast cancer tissues. *Biotechniques* 48: 389–397.
- Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, et al. (2005) Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR — how well do they correlate? *BMC Genomics* 6: 59.
- Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, et al. (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 98: 1183–1192.
- Baak JP, van Diest PJ, Voorhorst FJ, van der Wall E, Beex LV, et al. (2005) Prospective multicenter validation of the independent prognostic value of the mitotic activity index in lymph node-negative breast cancer patients younger than 55 years. *J Clin Oncol* 23: 5993–6001.
- Desmedt C, Sotiriou C (2006) Proliferation: the most prominent predictor of clinical outcome in breast cancer. *Cell Cycle* 5: 2198–2202.
- Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, et al. (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 10: R65.
- Starmans MH, Krishnapuram B, Steck H, Horlings H, Nuyten DS, et al. (2008) Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* 99: 1884–1890.
- Reyal F, van Vliet MH, Armstrong NJ, Horlings HM, de Visser KE, et al. (2008) A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. *Breast Cancer Res* 10: R93.
- Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365: 488–492.



47. Consortium M, Shi L, Reid LH, Jones WD, Shippy R, et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24: 1151–1161.
48. Weigelt B, Mackay A, A'Hern R, Natrajan R, Tan DS, et al. (2010) Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol* 11: 339–349.
49. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, et al. (2010) A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 16: 5222–5232.
50. Elloumi F, Hu Z, Li Y, Parker JS, Gulley ML, et al. (2011) Systematic bias in genomic classification due to contaminating non-neoplastic tissue in breast tumor samples. *BMC Med Genomics* 4: 54.
51. Venet D, Dumont JE, Detours V (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* 7: e1002240.
52. Rizzo P, Miao H, D'Souza G, Osipo C, Song LL, et al. (2008) Cross-talk between notch and the estrogen receptor in breast cancer suggests novel therapeutic approaches. *Cancer Res* 68: 5226–5235.
53. Simmons MJ, Serra R, Hermance N, Kelliher MA (2012) NOTCH1 inhibition in vivo results in mammary tumor regression and reduced mammary tumor-sphere forming activity in vitro. *Breast Cancer Res* 14: R126.
54. Dontu G, El-Ashry D, Wicha MS (2004) Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab* 15: 193–197.
55. Farnie G, Clarke RB (2007) Mammary stem cells and breast cancer—role of Notch signalling. *Stem Cell Rev* 3: 169–175.
56. Yao K, Rizzo P, Rajan P, Albain K, Rychlik K, et al. (2011) Notch-1 and notch-4 receptors as prognostic markers in breast cancer. *Int J Surg Pathol* 19: 607–613.
57. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, et al. (2010) Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer Res* 70: 709–718.
58. Haughian JM, Pinto MP, Harrell JC, Bliesner BS, Joensuu KM, et al. (2012) Maintenance of hormone responsiveness in luminal breast cancers by suppression of Notch. *Proc Natl Acad Sci U S A* 109: 2742–2747.