

**Keywords:** microarray; preoperative chemotherapy; gene expression profile; pathological complete response; disease-free survival

# Clinical and microarray analysis of breast cancers of all subtypes from two prospective preoperative chemotherapy studies

HS Okuma<sup>1,2</sup>, F Koizumi<sup>3</sup>, A Hirakawa<sup>4</sup>, M Nakatochi<sup>4</sup>, O Komori<sup>5,6</sup>, J Hashimoto<sup>1</sup>, M Kodaira<sup>1</sup>, M Yunokawa<sup>1</sup>, H Yamamoto<sup>1</sup>, K Yonemori<sup>1</sup>, C Shimizu<sup>1</sup>, Y Fujiwara<sup>1</sup> and K Tamura<sup>\*,1</sup>

<sup>1</sup>Department of Breast and Medical Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; <sup>2</sup>Department of Medical Oncology, Graduate School of Medicine, Chiba University Hospital, 1-8-1 Inohana, Chuo-ku, Chiba-shi, Chiba 260-8670, Japan; <sup>3</sup>Tokyo Metropolitan Cancer and Infectious diseases Center Komagome Hospital, 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8677, Japan; <sup>4</sup>Biostatistics and Bioinformatics Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan; <sup>5</sup>Department of Electrical and Electronics Engineering, University of Fukui, 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan and <sup>6</sup>The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan

**Background:** We aimed to analyse clinical and gene expression profiles to predict pathologic complete response and disease-free survival using two consecutive, prospective, preoperative chemotherapy trial cohorts.

**Methods:** Clinicopathological and gene expression data were evaluated in a cohort from two consecutive phase II preoperative studies that included patients with stage IIA–IIIC breast cancer of all subtypes. Analysed specimens were obtained before preoperative chemotherapy, and cDNA microarray analyses were performed using the Affymetrix Gene Chip U133 plus 2.0.

**Results:** Between December 2005 and December 2010, 122 patients were analysed. The pathologic complete response rate was significantly higher in HER2+ and HR –/HER2 – cancers. Age, pathologic complete response, HR –/HER2 – status, and lymph node positivity ( $\geq 4$ ) were significant poor prognostic factors for disease-free survival. For the cDNA microarray analyses, sufficient tumour samples were available from 78 of the 107 patients (73%). An 8-gene signature predictive of pathologic complete response and a 17-gene signature predictive of prognosis were identified. Patients were categorised into low-risk ( $n = 45$ ) and high-risk groups ( $n = 33$ ) (HR 70.0,  $P = 0.004$ ).

**Conclusions:** This study yielded preliminary data on the expression of specific genes predicting pathologic complete response and disease-free survival in a cohort of chemo-naïve breast cancer patients. Further validation may distinguish those who would benefit most from perioperative chemotherapy as well as those needing further intervention.

Breast cancer is the most common cancer in women worldwide. Today, the majority of patients with operable breast cancer are offered systemic therapy to reduce the risk of distant recurrence. Preoperative chemotherapy reduces the tumour burden of operable and locally advanced breast tumours (clinical stage IIIB–C) and thereby improves resectability, with the same survival effect as postoperative chemotherapy (van der Hage *et al*, 2001; Wolmark

*et al*, 2001; Rastogi *et al*, 2008; de Azambuja *et al*, 2014). In addition, preoperative treatment serves as an *in vivo* chemosensitivity test, allowing for early evaluation of the efficacy of chemotherapy (Schott and Hayes, 2012).

Clinical response rates to preoperative chemotherapy range from 60 to 80%, whereas pathologic complete response (pCR) rates are ~10–20% (Fisher *et al*, 1998; Smith *et al*, 2002), although both

\*Correspondence: Dr K Tamura; E-mail: ketamura@ncc.go.jp

Received 19 December 2015; revised 8 May 2016; accepted 19 May 2016; published online 14 July 2016

© 2016 Cancer Research UK. All rights reserved 0007 – 0920/16



differ among tumour subtypes. According to a previous report, the pCR rates for hormone receptor-positive (HR+)/human epidermal growth factor receptor 2-negative (HER2-), HR+/human epidermal growth factor receptor 2-positive (HER2+), hormone receptor-negative (HR-)/HER2+, and HR-/HER2- subtypes were 13%, 19%, 48%, and 29%, respectively (Iwata *et al*, 2011). Patients who achieve pCR have a better prognosis compared with those who do not (van der Hage *et al*, 2001; Wolmark *et al*, 2001; Rastogi *et al*, 2008). Studies investigating the factors associated with tumour response to chemotherapy have shown that markers of tumour cell proliferation, including Ki-67 staining, histologic grade, negative oestrogen receptor (ER) status, and HER2 overexpression, are significantly associated with the pCR rate (Petit *et al*, 2004; Dowsett *et al*, 2006; Andre *et al*, 2008; Nishimura *et al*, 2010). However, these results come with some degree of controversy and little indication of their clinically applicable predictive value (Burcombe *et al*, 2005). The biological mechanisms that influence tumour responsiveness in the pre-operative setting, including tumour recurrence, are not clearly understood.

Gene expression profiling studies in human tumours have provided new insights into the genes and pathways that contribute to tumorigenesis and the gene expression signatures that are prognostic of patient outcome. Previous studies aimed at discovering genes associated with breast cancer recurrence have uncovered several genes associated with cellular proliferation, resulting in the identification of genes associated with poor patient prognosis (Huang *et al*, 2003; Dai *et al*, 2005). These gene expression data have the potential to aid the determination of accurate, individualised prognosis, for example, through Agendia's MammaPrint and Genomic Health's 21-gene Oncotype Dx systems (van 't Veer *et al*, 2002; Paik *et al*, 2004; Paik *et al*, 2006). However, these approaches are based on node-negative, HR-positive early breast cancers, and no studies thus far have clearly identified gene signatures predicting both pCR and disease-free survival (DFS) from one cohort.

Here, we aimed to analyse the clinicopathological and gene expression profiles to predict pCR and DFS in early breast cancers managed by preoperative chemotherapy, using data from long-term follow-up of two prospective studies.

## MATERIALS AND METHODS

**Patients and samples.** We synthetically analysed the data of patients from our two previous consecutive prospective phase II studies in the preoperative chemotherapy setting, 'Trial A' (Tamura *et al*, 2011) and 'Trial B' (Ando *et al*, 2014). The two studies included patients with HER2+ and/or HER2-, and HR+ and/or HR- tumours. The trials had similar eligibility criteria; in brief, patients had histologically confirmed, previously untreated, unilateral, non-inflammatory invasive breast cancer. Histologic confirmation of invasive cancer was performed by core needle biopsy, and HER2- disease was assigned a score of 0 or 1+ based on immunohistochemistry or HER2 gene copy number: a chromosome 17 ratio of <2.0 by fluorescence *in situ* hybridisation (Wolff *et al*, 2007). Patients had clinical stage IIA-III primary measurable disease. Other requirements included age  $\geq 18$  years, Eastern Cooperative Oncology Group (ECOG) performance status (PS) score of 0-1, and adequate organ function (white blood cell count  $\geq 4000/\mu\text{l}$ , platelet count  $\geq 100\,000/\mu\text{l}$ , haemoglobin concentration  $\geq 9.0\text{ g dl}^{-1}$ , serum bilirubin  $\leq 2.0\text{ mg dl}^{-1}$ , aspartate aminotransferase and alanine aminotransferase  $\leq 100\text{ IU l}^{-1}$ , serum creatinine  $\leq$  institutional upper limit of normal range, PaO<sub>2</sub>  $\geq 60\text{ mm Hg}$ , and baseline left ventricular ejection fraction >50%).

Patients received 5-fluorouracil/epirubicin/cyclophosphamide (500/100/500 mg m<sup>-2</sup>) q3w  $\times 4$  cycles, followed by paclitaxel (PTX) (80 mg m<sup>-2</sup>) q1w  $\times 12$  cycles or docetaxel (75 mg m<sup>-2</sup>) q3w  $\times 4$  cycles or, if HER2+, PTX/trastuzumab (loading dose 4 mg kg<sup>-1</sup>, maintenance dose 2 mg kg<sup>-1</sup>) q1w  $\times 12$  cycles.

This study was conducted according to a protocol approved by the institutional review board and independent ethics committee, and informed consent was obtained from all patients for the use of biopsy specimens and the analysis of clinical information.

**Clinical statistical analysis.** Complete response rate was defined as no invasive residual tumour in breast and nodes, with noninvasive breast residuals allowed (ypT0/is, ypN0), a definition commonly used by MD Anderson Cancer Center, the Austrian Breast and Colorectal Cancer Study Group, and the Neo-Breast International Group (Green *et al*, 2005). Disease-free survival was estimated from the date of induction of adjuvant chemotherapy to the date of relapse or death from any cause (only relapses were considered events) using the Kaplan-Meier method. Potential predictive factors for pCR and recurrence were recorded, including patient age ( $\geq 35$  vs <35 years), tumour stage (IIa-IIb vs IIIa-IIIc), subtype (HR+/HER2- vs HR+/HER2+, HR-/HER2+, and HR-/HER2-), nuclear grade (1 vs 2-3), pCR rate (pCR vs non-pCR), and lymph node involvement. Nuclear grade, pCR, and lymph nodes were diagnosed with surgical specimens by two pathologists. A logistic regression model was used to estimate the odds ratio comparing the odds of pCR. Cox proportional hazards regression was used to investigate the prognostic factors for DFS.

**Microarray process and statistical analysis.** The mRNA was extracted from fine-needle biopsy samples performed at diagnosis before preoperative chemotherapy. cDNA was constructed from the mRNA of the breast cancer tissues by standard RT procedures. The probes were prepared by immunofluorescence labelling with cDNA and hybridised to chip arrays containing 54K probe sets (Affymetrix U133 plus 2.0). The data were processed by the robust multiarray average algorithm (Irizarry *et al*, 2003; Gautier *et al*, 2004) using GeneSpring ver. 12.6 (Agilent Technologies, Santa Clara, CA, USA). These processed values were corrected for batch effects using the ComBat function in the R package sva (Leek *et al*, 2012). Institution, age, stage, menopausal status, subtype, grade, ductal carcinoma *in situ*, pCR, and event DFS were used as covariates to correct for batch effects.

To identify a genomic signature of DFS using the data set after preprocessing, the 54 613 probes were filtered down to 104 probes, which satisfied the *P*-values of both the univariate Cox regression analysis and the significance analysis of microarrays (SAM) (Tusher *et al*, 2001) tests, which were <0.01. For the 104 probes, an iterative backward elimination feature selection procedure was applied using three-component partial Cox regression analysis (Li and Gui, 2004), where the partial Cox coefficients of the proteins were used for ranking (Ahdesmaki *et al*, 2013). During each iteration, the lowest ranking proteins were discarded after calculating the C-index (Harrell and Frank, 2010), and this was continued until the two probes remained. The C-index was calculated for each number of probes sets (i.e., from 104 to 2 of the length). In this study, we determined the length with highest C-index as the optimal one of the genomic signature.

Next, we developed the final genomic signature including these probes based on the partial Cox regression analysis. The DFS event risk based on the developed signature was quantified by the risk score that is defined as the linear combination of the signature probe values multiplied by their corresponding partial Cox model coefficients, first subtracted by the mean values of their probes.

That is, the risk score for patient can be written as

$$\text{Risk score} = \sum_{j=1}^L (\text{value of probe } j - \text{mean value of probe } j \text{ in } 78\text{pts})\beta_j,$$

where  $L$  is the number of optima length and  $\beta_j$  is the coefficient of the partial Cox regression model for probe  $j$ . The risk scores have by definition a sample mean of zero, and indicate high risk of DFS events for the larger value.

Finally, we calculate the internal accuracy of the final genomic signature. The survival curves for DFS between the two groups of the patients with  $\geq 0$  of risk score and patients with  $< 0$  were compared by the Kaplan–Meier method and log-rank test. The hazard ratio and its 95% confidence interval for the risk score dichotomised by zero were also estimated by the ordinary Cox regression analysis.

Additionally, we also identified the genes that were strongly correlated with pCR. Similar to the analyses for DFS, the 54 613 probes were filtered down to 363 probes that satisfied the  $P$ -values of both the univariate logistic regression analysis and the SAM tests, which were  $< 0.001$ . For the 363 probes, an iterative backward elimination feature selection procedure was applied using the ridge regression analysis (Friedman *et al*, 2010). Based on the area under the receiver operating characteristic curve (AUC), we determined the probes with highest AUC as the strongly correlated gene set.

All statistical tests were two sided. These analyses were performed using the SAS software (version 9.3; SAS Institute Inc, Cary, NC, USA) and the original R codes.

Average-linkage hierarchical clustering of genes and arrays was performed. We also conducted the clustering analysis using probes significant to the following gene groups: A, cell cycle regulator; B, signal; C, Rho family-related protein; D, blank; E, angiogenesis-related protein; F, growth factor; G, cytokine; H, apoptosis factors; I, DNA transcription factors/damage response, repair, and combination; J, metabolism/translation/protein turnover; K, detoxification enzyme; L, transporters/nucleocytoplasmic transporters/symporters and antiporters; M, cytoskeletal proteins; N, hormone-related and receptors; antibody-dependent cell-mediated cytotoxicity (ADCC), ADCC activators; Y, glyceraldehyde-3-phosphate dehydrogenase as a positive control.

Apart from the above-mentioned analysis, we re-identified the probes that correlated with pCR. To this end, we used the cohort from Trial A as a training set, and the cohort from Trial B as a validation set. Similar to the above-mentioned analysis, we first filtered the probes using the  $\chi^2$  test to remove probe sets that did not display significant variation in expression across arrays concerning pCR, a method used by the Cancer and Leukaemia Group B study for acute myeloid leukaemia (Marcucci *et al*, 2008). As the results, a total of  $\sim 8700$  probes met the filtering criteria and were included in the next step, where Wilcoxon's  $P$ -values by univariate and permutation tests were applied. The top-ranked probes were verified using the validation set based on the support vector machine with a linear kernel (Furey *et al*, 2000). The number of genes used in the formula was determined by a fivefold cross-validation. The HER2 status was also included in the formula.

## RESULTS

**Clinicopathological features predictive of pCR and DFS.** Between July 2007 and December 2010, 122 patients were enrolled in the two consecutive prospective studies. Of these patients, 107 underwent long-term follow-up study. The median follow-up time from the start of preoperative chemotherapy was 64.1 months (range 14.3–106 months). Table 1 summarises the patient

**Table 1. Baseline characteristics of patients from the two preoperative trials**

	Trial A (training set), N = 89		Trial B (validation set), N = 33	
	No.	%	No.	%
Median age (years), range	50 23–75		51 28–69	
<b>Menopausal status</b>				
Pre	30	34	12	36
Post	59	66	21	64
<b>ECOG performance status</b>				
0	83	93	32	97
1	6	6.7	1	3.1
<b>Stage</b>				
IIA	21	24	9	27
IIB	43	48	14	43
IIIA	15	17	5	15
IIIB	9	10	5	15
IIIC	1	1.1	0	0
<b>Histological grade</b>				
1	5	5.6	2	6
2	41	46	15	46
3	43	48	16	49
<b>No. of axillary lymph nodes</b>				
1	49	55	18	55
1–3	24	27	8	24
$\geq 4$	16	18	7	21
<b>Subtype</b>				
HR + /HER2 –	44	49	7	21
HR + /HER2 +	12	14	6	18
HR – /HER2 +	18	20	11	33
HR – /HER2 –	15	17	9	27
<b>Preoperative therapy</b>				
FEC × 4 + wPTX × 12	50	56	16	48
FEC × 4 + 3-wDTX × 4	8	9	0	0
FEC × 4 + wPTX + wT × 12	31	35	16	48
FEC × 4 + 3-wDTX × 4 + wT × 12	0	0	1	3
Pathologic complete response	26	29	12	36

Abbreviations: DTX = docetaxel; ECOG = Eastern Cooperative Oncology Group; FEC = fluorouracil + epirubicin + cyclophosphamide; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; PTX = paclitaxel; T = trastuzumab; w = weekly.

characteristics. The median age was 51 (23–75) years. The respective numbers of patients with ECOG PS of 0 and 1 were 115 and 7; with tumour stages IIA, IIB, IIIA, IIIB, and IIIC were 30, 57, 20, 14, and 1; and with tumour status HR + /HER2 –, HR + /HER2 +, HR – /HER2 +, and HR – /HER2 – were 51, 18, 29, and 24. The characteristics and results of the 107 patients were comparable to those of the original cohort (122 patients). The pCR rate of the 107 patients was 28% (HR + /HER2 – : HR + /HER2 + : HR – /HER2 + : HR – /HER2 – = 2.1% : 38% : 70% : 35%).

Table 2A depicts the multivariate analysis with clinicopathological characteristics of predictive factors for pCR. The pCR rate was significantly higher in HR + /HER2 +, HR – /HER2 +, and HR – /HER2 – breast cancers compared with HR + /HER2 – cancers ( $P = 0.004$ ,  $P < 0.001$ , and  $P = 0.007$ , respectively). The 5-year DFS for all subtypes was 77.4%. The survival curves for each subtype are shown in Figure 1. The HR – /HER2 – subtype had a significantly poor prognosis ( $P = 0.0045$ ). Multivariate analysis revealed that non-pCR, age  $< 35$  years, HR – /HER2 – subtype, and axillary lymph nodes  $\geq 4$  were significant poor prognostic factors ( $P = 0.03$ ,  $P = 0.005$ ,  $P < 0.001$ , and  $P < 0.001$ , respectively), as shown in Table 2B.

**Table 2A. Factors predictive of pCR assessed by multivariate logistic regression model (n = 107)**

	Odds ratio	95% CI	P-value
<b>Age (years)</b>			
≥ 35	1		
< 35	1.22	(0.19, 7.79)	0.83
<b>Stage</b>			
IIA/IIB	1		
IIIA/IIIB/IIIC	0.4	(0.12, 1.34)	0.14
<b>Subtype</b>			
HR + /HER2 -	1		
HR + /HER2 +	38.7	(3.27, 458)	0.004
HR - /HER2 +	142	(13.1, 1540)	<0.001
HR - /HER2 -	26.3	(2.45, 282)	0.007
<b>Histological grade</b>			
1	1		
2 or 3	0.55	(0.07, 4.06)	0.56

Abbreviations: CI = confidence interval; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; pCR = pathologic complete response.

**Table 2B. Factors predictive of DFS assessed by multivariate Cox regression model (n = 107)**

	Hazard ratio	95% CI	P-value
<b>Response</b>			
Non-pCR	1		
pCR	0.09	(0.09, 0.81)	0.03
<b>Age (years)</b>			
≥ 35	1		
< 35	7.76	(1.83, 32.9)	0.005
<b>Stage</b>			
IIA/IIB	1		
IIIA/IIIB/IIIC	1.17	(0.45, 3.02)	0.75
<b>Subtype</b>			
HR + /HER2 -	1		
HR + /HER2 +	2.28	(0.39, 13.4)	0.36
HR - /HER2 +	3.7	(0.87, 15.8)	0.08
HR - /HER2 -	8.09	(2.54, 25.8)	<0.001
<b>Histological grade</b>			
1	1		
2 or 3	0.16	(0.02, 1.17)	0.07
<b>No. of axillary lymph nodes</b>			
0	1		
1-3	3.19	(0.89, 11.4)	0.07
≥ 4	12.3	(3.33, 45.6)	<0.001

Abbreviations: CI = confidence interval; DFS = disease-free survival; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; pCR = pathologic complete response.

**Microarray analysis.** Primary breast cancer tissues were obtained from the 107 patients by fine-needle biopsy; 78 (73%) samples contained sufficient mRNA for cDNA microarray analysis. Results of the gene signature analysis identified an 8-gene expression profile predictive of pCR and a 17-gene expression profile most associated with DFS, as shown in Table 3A and B. With the DFS 17-gene signature, which includes the apoptosis-related cysteine peptidase caspase-8 encoding gene (*CASP8*), patients were classified into a low-risk group (n = 45) and a high-risk group (n = 33) according to the risk score determined by the partial Cox regression model (Figure 2). A Cox proportional hazards regression to investigate the prognostic factors, including the gene expression profiles of low- and high-risk factors, for DFS showed that the gene profiling was the strongest factor (Table 4).

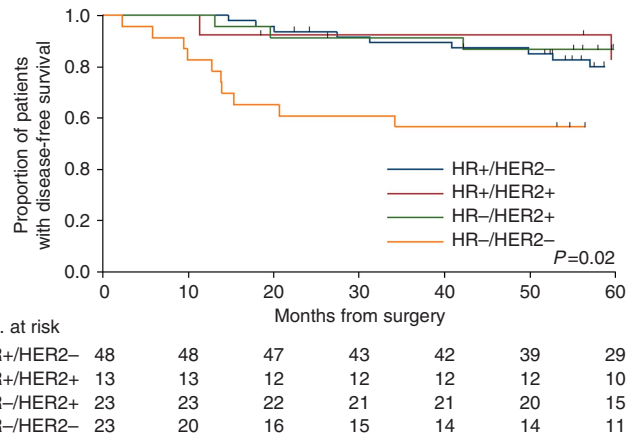


Figure 1. Kaplan-Meier disease-free survival curves of patients stratified by subtype (n = 107). The median follow-up time was 64 months. The 5-year DFS% (95% confidence interval (CI)) were as follows: HR + /HER2 - , 79.9% (64.8, 89.0); HR + /HER2 + , 83.1% (47.2, 95.5); HR - /HER2 + , 86.7% (64.3, 95.5); and HR - /HER2 - , 56.5% (34.3, 73.8).

Based on gene clustering analysis, the breast cancers were classified according to the following immunohistochemistry subtypes: luminal A and B, HER2-enriched, and triple-negative breast cancer (TNBC). Triple-negative breast cancer was further divided into two clusters most likely indicating basal-like and claudin-low intrinsic subtypes, as previously reported by Sorlie *et al* (2001), indicating the high quality of the representative method (Supplementary Data and Figure 1). The genes used in the clustering analysis is shown in Supplementary Table 1. When the clustering analysis using significant gene groups of A to N and Y was performed according to HER2 status, HER2 + breast cancers were successfully clustered using the B group gene probes, whereas HER2 - breast cancers were clustered appropriately with the gene probes of the N, J, I, and H groups (data not shown).

In addition, using the training set, we identified the top-ranked gene probes associated with pCR (Wilcoxon's P-value < 0.01). On the basis of the 100 top-ranked genes (Supplementary Table 2), the training set was also used for model selection and resulted in the creation of a gene model for the HER2 - breast cancers (HR + /HER2 - and HR - /HER2 -) (Supplementary Figure 2). This analysis included 49 samples, which included 35 from the training set and 14 from the validation set. The discrimination markers predicting pCR were analysed using the training and validation sets, and the lowest test error of 7% was observed when the number of markers fitted in the model was four; these included HER2 - status and the following three genes: promyelocytic leukaemia protein (*PML*), tryptophanyl-tRNA synthetase (*TrpRS*), and choline kinase alpha (*CHKA*) (Supplementary Figure 3). A similar analysis including the HER2 + subset failed to show a well-fitted model.

**DISCUSSION**

Breast cancer is heterogeneous in terms of prognosis and response to chemotherapy, even among known intrinsic subtypes. Results from our clinicopathological data revealed that HER2 + and HR - /HER2 - subtypes were predictive of pCR. A significantly poor DFS was observed in the HR - /HER2 - subtype and in lymph node-positive tumours. These results are comparable with results previously reported such as a study by Liedtke *et al* (2008); therefore, our cohort of two combined trials is representative of the



**Table 3A. Eight genes contained in the pCR discriminating profile (n = 78)**

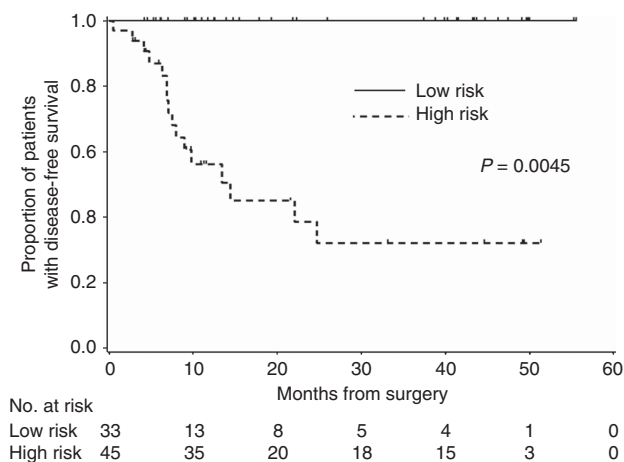
Gene bank ID	Gene symbol	Gene title	Odds ratio	Odds ratio, 95% CI
1553973_a_at	SPINK6	Serine peptidase inhibitor, Kazal type 6	0.00001	0–0.04
1554666_at	LOC100130950	Uncharacterized LOC100130950	0.00001	0–0.05
1558392_at	SYNE2	Spectrin repeat containing, nuclear envelope 2	134	11.8–1528
1564707_x_at	GLS2	Glutaminase 2 (liver, mitochondrial)	0.00001	0–0.01
1570541_s_at	GBP1P1	Guanylate binding protein 1, interferon-inducible pseudogene 1	105	6.87–1611
205746_s_at	ADAM17	ADAM metallopeptidase domain 17	172	9.51–3121
206478_at	KIAA0125	KIAA0125	0.00001	0–0.07
208557_at	HOXA6	Homeobox A6	692	25.1–19038

Abbreviations: CI = confidence interval; pCR = pathologic complete response.

**Table 3B. Seventeen genes contained in the DFS discriminating profile (n = 78)**

Gene bank ID	Gene symbol	Gene title	Hazard ratio	Hazard ratio, 95% CI
1556236_at	—	—	5.49	2.38–12.64
201956_s_at	GNPAT	Glyceronephosphate O-acyltransferase	0.07	0.01–0.31
202321_at	GGPS1	Geranylgeranyl diphosphate synthase 1	0.16	0.06–0.46
203283_s_at	HS2ST1	Heparan sulfate 2-O-sulfotransferase 1	6.47	2.83–14.82
207254_at	SLC15A1	Solute carrier family 15 (oligopeptide transporter), member 1	3.72	2.01–6.9
210861_s_at	WISP3	WNT1-inducible signalling pathway protein 3	2.2	1.43–3.39
212479_s_at	RMND5A	Required for meiotic nuclear division 5 homolog A ( <i>S. cerevisiae</i> )	8.54	2.69–27.16
213373_s_at	CASP8	Caspase 8, apoptosis-related cysteine peptidase	0.22	0.1–0.49
218783_at	INTS7	Integrator complex subunit 7	0.22	0.09–0.5
219958_at	TMEM74B	Transmembrane protein 74B	2.5	1.54–4.05
228542_at	MRS2	MRS2 magnesium transporter	5.29	2.21–12.66
230769_at	DENND2C	DENN/MADD domain containing 2C	3.72	1.71–8.07
231292_at	EID3	EP300 interacting inhibitor of differentiation 3	6.28	2.25–17.54
234726_s_at	TMEM168	Transmembrane protein 168	9.64	2.87–32.34
235940_at	C9orf64	Chromosome 9 open reading frame 64	0.32	0.18–0.56
239282_at	CCDC41	Coiled-coil domain containing 41	8.75	2.82–27.1
241721_at	—	—	0.11	0.04–0.28

Abbreviations: CI = confidence interval; DFS, disease-free survival.



**Figure 2.** Patients were classified into low-risk (n = 45, bold line) and high-risk (n = 33, dotted line) groups with the 17-gene signature predicting DFS by partial least-squares Cox regression. The hazard ratio was 67.8 (95% confidence interval (CI), 3.70–1240), P = 0.0045.

general cancer population. Liedtke *et al* (2008) found that the TNBC subtype had a higher pCR rate than the non-TNBC subtype, but a significantly decreased survival. This was explained by the heterogeneity of the TNBC subtype, as some patients achieved pCR with a good prognosis, whereas the majority did not achieve pCR and had a significantly worse prognosis than other subtypes. Lehmann *et al* (2011) proposed that TNBC could be classified into seven molecular subtypes, among which the pCR rate differed, as shown by Masuda *et al* (2013). These data, together with our own, suggest heterogeneity even among the known breast cancer subtypes. The finding that four or more positive axillary lymph nodes constituted a poor prognostic factor correlates with historical data showing that lymph node positivity is the most established and reliable prognostic factor for subsequent metastatic disease and survival (Fisher *et al*, 1993).

The microarray analysis that included the cohorts from two prospective preoperative studies yielded preliminary data showing that an 8-gene signature predicted pCR and a 17-gene signature predicted DFS. With the 17-gene signature we were able to discriminate low- and high-risk patients with a high hazard ratio of 70, showing a higher discriminating value than single clinical

**Table 4. Factors predictive of DFS assessed by multivariate Cox regression model in patients from gene analysis cohort (n = 78)**

	Hazard ratio	95% CI	P-value
<b>Gene expression risk score</b>			
Low	1		
High	6.68	(0.94, 47.1)	0.05
<b>Response</b>			
Non-pCR	1		
pCR	0.22	(0.01, 3.87)	0.3
<b>Age (years)</b>			
≥50	1		
<50	1.8	(0.26, 12.4)	0.55
<b>Subtype</b>			
HR + /HER2 –	1		
HR + /HER2 +	1.14	(0.10, 12.3)	0.9
HR – /HER2 +	1.37	(0.09, 20.9)	0.81
HR – /HER2 –	1.02	(0.08, 12.2)	0.98
<b>Histological grade</b>			
1 or 2	1		
3	0.67	(0.10, 4.48)	0.68
<b>No. of axillary lymph nodes</b>			
0	1		
1–3	0.69	(0.09, 5.25)	0.72
≥4	1.02	(0.17, 5.87)	0.97

Abbreviations: CI = confidence interval; DFS = disease-free survival; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; pCR = pathologic complete response.

variables (e.g., histologic grade, HR status, HER2 status). Many studies have aimed to identify gene profiles predicting response and prognosis, including commercially available panels (van't Veer *et al*, 2002; Paik *et al*, 2004; Paik *et al*, 2006; Nishio *et al*, 2014). Most are based on node-negative early breast cancers, and include only HR – or HR + cancers. This study combined all subtypes, and identified two sets of gene clusters that, respectively, predicted pCR and PFS. By including all subtypes in the analysis, we tried to seek for a molecular feature that explains the difference in chemosensitivity and prognosis beyond the known biomarkers. There is more hidden biology than the already defined subtypes using HER2 and hormone receptor proteins, as the prognosis differs even among the same subtypes (Liedtke *et al*, 2008). Also, discordance of subtype between primary and metastatic sites have been reported (Falck *et al*, 2013; Yao *et al*, 2014), which also implies that the cancer is driven beyond the known biomarkers. Most molecular analysis in the literature are carried out in breast cancers of a certain HER2 or HR status and therefore our analysis is novel in that the identified signatures may be useful regardless of subtype and those with discordances or changes of subtype at metastatic sites or at relapse.

The 8-gene signature may identify which patients will benefit most from preoperative chemotherapy and which should initially undergo surgery. With the DFS prediction signature, we may be able to predict those who will benefit from additional adjuvant chemotherapy or extended hormone therapy.

The genes that were chosen in our two signatures did not show an overlap with genes in other microarray studies, namely those assessed with the MammaPrint (Agendia, Irvine, CA, USA) and Oncotype DX (Genomic Health, Redwood City, CA, USA) panels. Possible explanations are that our study included all subtypes regardless of node status, as well as patients who were treated or not treated with trastuzumab. Nonetheless, we identified a gene signature with a high prognostic power, assuming that molecular pathways other than the already known molecular subtypes are associated with tumour recurrence. Further, in our study the gene

signature predicting pCR and DFS did not identify any overlapping genes between the two signatures. This result is supported by our clinicopathological data sets, where a discordance between pCR and DFS also occurred among subtypes. The HR – /HER2 – subtype was associated with a significantly poor DFS compared with other subtypes, despite a relatively high pCR rate, which was also shown by Liedtke *et al* (2008). The NeoALTTO trial showed that dual anti-HER2 inhibition significantly increased pCR rate compared with trastuzumab alone, but at the same time did not show a survival difference between treatment groups (de Azambuja *et al*, 2014). A recent meta-analysis concluded that the pCR rate should not be used as a surrogate marker for survival (Cortazar *et al*, 2014). Therefore, response to chemotherapy and prognosis outcome may be a result from different biology. Using the two different signatures for analysis of tumour response to preoperative chemotherapy and its prognosis may provide a new means of making individual therapeutic decisions.

Further, discordance of genomic signatures among studies is a well-discussed topic, and it is known that independent signatures may be similar in terms of outcome predictions despite a lack of gene overlap. Our Supplementary Data analysis in the HER2 – subtype, resulting in a 3-gene signature, also did not show an overlap with the 8- or 17-gene signature. One explanation is that the number of samples used were different (49 vs 78). Also, two different analysis methods were used. The results of the Supplementary Data used a svm with linear kernel, which mainly focuses on the error rate, rather than the odds ratio.

Our 8-gene signature predicting pCR included genes associated with tumorigenesis such as ADAM metalloproteinase domain 17 (ADAM17), glutaminase-2 (GLS2), and HOXA6. The ADAM17 protein catalyses the release and activation of ligands such as transforming growth factor- $\alpha$ , which is essential in activating epidermal growth factor receptor (Blobel, 2005), and high expression of ADAM17 was associated with shorter survival in breast cancers (McGowan *et al*, 2008). Recently, *in vitro* data suggested that an antibody against ADAM17 had antitumour effects in TNBC cells (Caiazza *et al*, 2015). Glutaminase is involved in the Warburg effect in cancer cells, and two human glutaminase genes, GLS1 and GLS2, have been identified (Erickson and Cerione, 2010). A discovery of an alkyl benzoquinones that preferentially inhibit GLS2 and subsequently reduces carcinoma cell proliferation and induced autophagy via AMPK-mediated mTORC1 inhibition has been reported (Lee *et al*, 2014). It suggested that the inhibition of GLS2 may be a potential anticancer target. HOX genes encode a highly conserved family of homeodomain-containing transcription factors that have crucial roles in determining the identity of cells and tissues during embryogenesis. Aberrant HOX gene expression has been linked to a variety of adult malignancies (Shah and Sukumar, 2010). In a study to analyse the expression of 39 HOX genes in malignant breast tissues, HOXA6 showed low expression levels along with HOXB8 and HOXC5 in malignant tissues, whereas the other HOX family genes were expressed at higher levels (Hur *et al*, 2014). Our 17-gene signature that predicted DFS included CASP8, an apoptosis-related cysteine peptidase encoding gene. Caspase-8 has a central role in the transmission of the death signal in the death receptor (extrinsic) pathway of apoptosis by coupling the stimulation of death receptors to the activation of intracellular signalling cascades that eventually lead to cell death (Barnhart *et al*, 2003). It is frequently inactivated in tumours of breast, colon, or lung (Shivapurkar *et al*, 2002). Therapeutic agents such as interferon- $\gamma$  and peptides induce caspase pathways, and microtubule-stabilising agents such as taxanes promote CASP8-mediated apoptosis, possibly via upregulation of components of the tumour necrosis factor-related apoptosis-inducing ligand pathway (Nimmanapalli *et al*, 2001; Muhlethaler-Mottet *et al*, 2004) or by amplification of CASP8 activation via microtubule-anchored death effector domain 'filaments' (Mielgo *et al*, 2009). Based on these

data, we assume that patients classified into the high-risk group by our 17-gene signature may benefit from more intensive chemotherapy, and those in the low-risk group may be able to avoid extra chemotherapy.

As for the Supplementary Data, analysis using the training and validation sets identified HER2 – status as well as three genes, namely *PML*, *TrpRS*, and *CHKA*, as the most accurate markers for predicting pCR. *PML* is a tumour suppressor gene as its expression is lost in several types of human cancers and is associated with tumour grade and progression (Gurrieri *et al*, 2004). *TrpRS* has been documented to function in proangiogenic responses (Wakasugi and Schimmel, 1999; Mirando *et al*, 2014). Human *TrpRS* exists in two forms: full-length protein and truncated *TrpRS*. The expression of mini *TrpRS* is stimulated by the antitumorigenic *IGN- $\gamma$*  (Rubin *et al*, 1991), and in a protein signature analysis with TNBC, *TrpRS* was identified as a good prognostic marker (Campone *et al*, 2015). *CHKA* catalyses the phosphorylation of choline and has been shown to be upregulated in many cancer types, including breast, lung, colorectal, and prostate cancer (Katz-Brull *et al*, 2002). When the gene is amplified, it modulates ER-driven proliferation and ER/oestrogen response element transactivation (Lopez-Knowles *et al*, 2015). With this three-gene model, validation set was used for performance evaluation, which resulted in a 93% accuracy rate on validation. The clustering analysis revealed that the HER2+ cancers correlated with only the B group gene probes, whereas the HER2 – cancers showed a correlation with the N, J, I, and H groups. According to a pooled analysis of studies of gene modules and response to preoperative chemotherapy (Ignatiadis *et al*, 2012), high module scores for chromosomal instability, phosphatase and tensin homolog loss, and E2F3 transcription factor were associated with increased pCR probability in HER2 – tumours. These findings, along with our current results, suggest that HER2 – cancers possess marked heterogeneity, with many different processes and pathways associated with tumour growth or sensitivity to chemotherapy. This is in contrast to HER2+ cancers, in which HER2 amplification-driven oncogenesis has a much more significant role than other pathways. Therefore, we believe that using the proper sets of genes, pCR could be more successfully predicted in HER2 – cancers compared with those that are HER2+, would be most successful in predicting pCR with the proper sets of genes.

Some limitations of our analysis must be mentioned. First, the relatively small sample size may have increased the false discovery-positive rate. Second, all patients in the analysis received anthracycline plus taxane-based preoperative chemotherapy regimens. It is therefore not known if the associations between significant genes and pCR or DFS are anthracycline-specific or taxane-specific, or instead indicate general chemosensitivity. Third, validation of the gene signatures was not performed owing to the lack of events, which will be conducted in the future. Fourth, our analysis integrated all subtypes, which is also the novelty of our analysis, but identifying a predictive gene signature for each subtype may have been more suitable for clinical use. However, if analysis was carried out for each subtype, the number of cases would be limited with a decrease in the power of the test, leading to a less meaningful gene signature. We believe that enrolling patients in a prospective trial, as performed in our study, should decrease the number of accidental findings related to associated genes. Furthermore, we performed an integrated analysis of the two trials, using them as a training set and validation set, respectively.

Our clinicopathological and gene expression analyses for prediction of response to chemotherapy and recurrence resulted in two preliminary set of genes predictive of pCR and DFS, which may provide guidance regarding individual therapeutic decisions. RNA genomic analysis was feasible in 73% of the specimens, resulting in successful molecular classification, suggesting that

HER2 – tumours have a higher likelihood of gene clustering. We wish to further expand and validate this analysis by using integrated data from larger retrospective populations from previously published studies (Madden *et al*, 2013) or global prospective trials.

## ACKNOWLEDGEMENTS

We thank all the study participants who provided tissue samples for this analysis. We also thank Ms Y Kitamura from the Shien-Laboratory. Finally, we thank Ms R Koyama for secretarial support in this study. This study was funded by a grant from the Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labour, and Welfare of Japan (H22-general-23).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- Ahdesmaki M, Lancashire L, Proutski V, Wilson C, Davison TS, Harkin DP, Kennedy RD (2013) Model selection for prognostic time-to-event gene signature discovery with applications in early breast cancer data. *Stat Appl Genet Mol Biol* **12**: 619–635.
- Ando M, Yamauchi H, Aogi K, Shimizu S, Iwata H, Masuda N, Yamamoto N, Inoue K, Ohono S, Kuroi K, Hamano T, Sukigara T, Fujiwara Y (2014) Randomized phase II study of weekly paclitaxel with and without carboplatin followed by cyclophosphamide/epirubicin/5-fluorouracil as neoadjuvant chemotherapy for stage II/IIIA breast cancer without HER2 overexpression. *Breast Cancer Res Treat* **145**: 401–409.
- Andre F, Mazouni C, Liedtke C, Kau SW, Frye D, Green M, Gonzalez-Angulo AM, Symmans WF, Hortobagyi GN, Pusztai L (2008) HER2 expression and efficacy of preoperative paclitaxel/FAC chemotherapy in breast cancer. *Breast Cancer Res Treat* **108**: 183–190.
- Barnhart BC, Lee JC, Alappat EC, Peter ME (2003) The death effector domain protein family. *Oncogene* **22**: 8634–8644.
- Blobel CP (2005) ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol* **6**: 32–43.
- Burcombe RJ, Makris A, Richman PI, Daley FM, Noble S, Pittam M, Wright D, Allen SA, Dove J, Wilson GD (2005) Evaluation of ER, PgR, HER-2 and Ki-67 as predictors of response to neoadjuvant anthracycline chemotherapy for operable breast cancer. *Br J Cancer* **92**: 147–155.
- Caiazza F, McGowan PM, Mullyooly M, Murray A, Synnott N, O'Donovan N, Flanagan L, Tape CJ, Murphy G, Crown J, Duffy MJ (2015) Targeting ADAM-17 with an inhibitory monoclonal antibody has antitumour effects in triple-negative breast cancer cells. *Br J Cancer* **112**: 1895–1903.
- Campone M, Valo I, Jezequel P, Moreau M, Boissard A, Champion L, Loussouarn D, Verrielle V, Coqueret O, Guette C (2015) Prediction of recurrence and survival for triple-negative breast cancer (TNBC) by a protein signature in tissue samples. *Mol Cell Proteomics* **14**: 2936–2946.
- Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, Swain SM, Prowell T, Loibl S, Wickerham DL, Bogaerts J, Baselga J, PEROU C, Blumenthal G, Blohmer J, Mamounas EP, Bergh J, Semiglazov V, Justice R, Eidtmann H, Paik S, Piccart M, Sridhara R, Fasching PA, Slaets L, Tang S, Gerber B, Geyer Jr CE, Pazdur R, Ditsch N, Rastogi P, Eiermann W, von Minckwitz G (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* **384**: 164–172.
- Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, Bernards R, van de Vijver M, Deutsch P, Sachs A, Stoughton R, Friend S (2005) A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* **65**: 4059–4066.
- de Azambuja E, Holmes AP, Piccart-Gebhart M, Holmes E, di Cosimo S, Swaby RF, Untch M, Jackisch C, Lang I, Smith I, Boyle F, Xu B, Barrios CH, Perez EA, Azim Jr HA, Kim SB, Kuemmel S, Huang CS,



- Vuyksteke P, Hsieh RK, Gorbunova V, Eniu A, Dreosti L, Tavartkiladze N, Gelber RD, Eidtmann H, Baselga J (2014) Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): survival outcomes of a randomised, open-label, multicentre, phase 3 trial and their association with pathological complete response. *Lancet Oncol* **15**: 1137–1146.
- Dowsett M, Smith IE, Ebbs SR, Dixon JM, Skene A, Griffith C, Boeddinghaus I, Salter J, Detre S, Hills M, Ashley S, Francis S, Walsh G, A'Hern R (2006) Proliferation and apoptosis as markers of benefit in neoadjuvant endocrine therapy of breast cancer. *Clin Cancer Res* **12**: 1024s–1030s.
- Erickson JW, Cerione RA (2010) Glutaminase: a hot spot for regulation of cancer cell metabolism? *Oncotarget* **1**: 734–740.
- Falck AK, Bendahl PO, Chebil G, Olsson H, Ferno M, Ryden L (2013) Biomarker expression and St Gallen molecular subtype classification in primary tumours, synchronous lymph node metastases and asynchronous relapses in primary breast cancer patients with 10 years' follow-up. *Breast Cancer Res Treat* **140**: 93–104.
- Fisher B, Bryant J, Wolmark N, Mamounas E, Brown A, Fisher ER, Wickerham DL, Begovic M, Decillis A, Robidoux A, Margolese RG, Cruz Jr AB, Hoehn JL, Lees AW, Dimitrov NV, Bear HD (1998) Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol* **16**: 2672–2685.
- Fisher ER, Anderson S, Redmond C, Fisher B (1993) Pathologic findings from the National Surgical Adjuvant Breast Project protocol B-06. 10-year pathologic and clinical prognostic discriminants. *Cancer* **71**: 2507–2514.
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* **33**: 1–22.
- Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**: 906–914.
- Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy – analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307–315.
- Green MC, Buzdar AU, Smith T, Ibrahim NK, Valero V, Rosales MF, Cristofanilli M, Booser DJ, Pusztai L, Rivera E, Theriault RL, Carter C, Frye D, Hunt KK, Symmans WF, Strom EA, Sahin AA, Sikov W, Hortobagyi GN (2005) Weekly paclitaxel improves pathologic complete remission in operable breast cancer when compared with paclitaxel once every 3 weeks. *J Clin Oncol* **23**: 5983–5992.
- Gurrieri C, Capodici P, Bernardi R, Scaglioni PP, Nafa K, Rush LJ, Verbel DA, Cordon-Cardo C, Pandolfi PP (2004) Loss of the tumor suppressor PML in human cancers of multiple histologic origins. *J Natl Cancer Inst* **96**: 269–279.
- Harrell J, Frank E (2010) *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer: New York, NY, USA.
- Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR, Huang AT (2003) Gene expression predictors of breast cancer outcomes. *Lancet* **361**: 1590–1596.
- Hur H, Lee JY, Yun HJ, Park BW, Kim MH (2014) Analysis of HOX gene expression patterns in human breast cancer. *Mol Biotechnol* **56**: 64–71.
- Ignatiadis M, Singhal SK, Desmedt C, Haibe-Kains B, Criscitiello C, Andre F, Loi S, Piccart M, Michiels S, Sotiriou C (2012) Gene modules and response to neoadjuvant chemotherapy in breast cancer subtypes: a pooled analysis. *J Clin Oncol* **30**: 1996–2004.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.
- Iwata H, Sato N, Masuda N, Nakamura S, Yamamoto N, Kuroi K, Kurosumi M, Tsuda H, Akiyama F, Ohashi Y, Toi M (2011) Docetaxel followed by fluorouracil/epirubicin/cyclophosphamide as neoadjuvant chemotherapy for patients with primary breast cancer. *Jpn J Clin Oncol* **41**: 867–875.
- Katz-Brull R, Seger D, Rivenson-Segal D, Rushkin E, Degani H (2002) Metabolic markers of breast cancer: enhanced choline metabolism and reduced choline-ether-phospholipid synthesis. *Cancer Res* **62**: 1966–1970.
- Lee YZ, Yang CW, Chang HY, Hsu HY, Chen IS, Chang HS, Lee CH, Lee JC, Kumar CR, Qiu YQ, Chao YS, Lee SJ (2014) Discovery of selective inhibitors of glutaminase-2, which inhibit mTORC1, activate autophagy and inhibit proliferation in cancer cells. *Oncotarget* **5**: 6087–6101.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**: 882–883.
- Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**: 2750–2767.
- Li H, Gui J (2004) Partial Cox regression analysis for high-dimensional microarray gene expression data. *Bioinformatics* **20**(Suppl 1): i208–i215.
- Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L (2008) Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* **26**: 1275–1281.
- Lopez-Knowles E, Wilkerson PM, Ribas R, Anderson H, Mackay A, Ghazoui Z, Rani A, Osin P, Nerurkar A, Renshaw L, Larionov A, Miller WR, Dixon JM, Reis-Filho JS, Dumbier AK, Martin LA, Dowsett M (2015) Integrative analyses identify modulators of response to neoadjuvant aromatase inhibitors in patients with early breast cancer. *Breast Cancer Res* **17**: 35.
- Madden SF, Clarke C, Gaule P, Aherne ST, O'Donovan N, Clynes M, Crown J, Gallagher WM (2013) BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. *Breast Cancer Res* **15**: R52.
- Marcucci G, Radmacher MD, Maharry K, Mrozek K, Ruppert AS, Paschka P, Vukosavljevic T, Whitman SP, Baldus CD, Langer C, Liu CG, Carroll AJ, Powell BL, Garzon R, Croce CM, Kolitz JE, Caligiuri MA, Larson RA, Bloomfield CD (2008) MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med* **358**: 1919–1928.
- Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, Valero V, Lehmann BD, Pietenpol JA, Hortobagyi GN, Symmans WF, Ueno NT (2013) Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin Cancer Res* **19**: 5533–5540.
- McGowan PM, Mckiernan E, Bolster F, Ryan BM, Hill AD, Mcdermott EW, Evoy D, O'Higgins N, Crown J, Duffy MJ (2008) ADAM-17 predicts adverse outcome in patients with breast cancer. *Ann Oncol* **19**: 1075–1081.
- Mielgo A, Torres VA, Clair K, Barbero S, Stupack DG (2009) Paclitaxel promotes a caspase 8-mediated apoptosis through death effector domain association with microtubules. *Oncogene* **28**: 3551–3562.
- Mirando AC, Francklyn CS, Lounsbury KM (2014) Regulation of angiogenesis by aminoacyl-tRNA synthetases. *Int J Mol Sci* **15**: 23725–23748.
- Muhlethaler-Mottet A, Boursouf KB, Auderset K, Joseph JM, Gross N (2004) Drug-mediated sensitization to TRAIL-induced apoptosis in caspase-8-complemented neuroblastoma cells proceeds via activation of intrinsic and extrinsic pathways and caspase-dependent cleavage of XIAP, Bcl-xL and RIP. *Oncogene* **23**: 5415–5425.
- Nimmanapalli R, Perkins CL, Orlando M, O'Bryan E, Nguyen D, Bhalla KN (2001) Pretreatment with paclitaxel enhances apo-2 ligand/tumor necrosis factor-related apoptosis-inducing ligand-induced apoptosis of prostate cancer cells by inducing death receptors 4 and 5 protein levels. *Cancer Res* **61**: 759–763.
- Nishimura R, Osako T, Okumura Y, Hayashi M, Arima N (2010) Clinical significance of Ki-67 in neoadjuvant chemotherapy for primary breast cancer as a predictor for chemosensitivity and for prognosis. *Breast Cancer* **17**: 269–275.
- Nishio M, Naoi Y, Tsunashima R, Nakauchi C, Kagara N, Shimoda M, Shimomura A, Maruyama N, Shimazu K, Kim SJ, Noguchi S (2014) 72-gene classifier for predicting prognosis of estrogen receptor-positive and node-negative breast cancer patients using formalin-fixed, paraffin-embedded tumor tissues. *Clin Breast Cancer* **14**: e73–e80.
- 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 (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, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer Jr CE, Wickerham DL, Wolmark N (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24**: 3726–3734.
- Petit T, Wilt M, Velten M, Millon R, Rodier JF, Borel C, Mors R, Haegle P, Eber M, Ghnassia JP (2004) Comparative value of tumour grade, hormonal receptors, Ki-67, HER-2 and topoisomerase II alpha status as



- predictive markers in breast cancer patients treated with neoadjuvant anthracycline-based chemotherapy. *Eur J Cancer* **40**: 205–211.
- Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, Margolese RG, Hoehn JL, Vogel VG, Dakhil SR, Tamkus D, King KM, Pajon ER, Wright MJ, Robert J, Paik S, Mamounas EP, Wolmark N (2008) Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. *J Clin Oncol* **26**: 778–785.
- Rubin BY, Anderson SL, Xing L, Powell RJ, Tate WP (1991) Interferon induces tryptophanyl-tRNA synthetase expression in human fibroblasts. *J Biol Chem* **266**: 24245–24248.
- Schott AF, Hayes DF (2012) Defining the benefits of neoadjuvant chemotherapy for breast cancer. *J Clin Oncol* **30**: 1747–1749.
- Shah N, Sukumar S (2010) The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* **10**: 361–371.
- Shivapurkar N, Reddy J, Matta H, Sathyanarayana UG, Huang CX, Toyooka S, Minna JD, Chaudhary PM, Gazdar AF (2002) Loss of expression of death-inducing signaling complex (DISC) components in lung cancer cell lines and the influence of MYC amplification. *Oncogene* **21**: 8510–8514.
- Smith IC, Heys SD, Hutcheon AW, Miller ID, Payne S, Gilbert FJ, Ah-See AK, Eremin O, Walker LG, Sarkar TK, Eggleton SP, Ogston KN (2002) Neoadjuvant chemotherapy in breast cancer: significantly enhanced response with docetaxel. *J Clin Oncol* **20**: 1456–1466.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* **98**: 10869–10874.
- Tamura K, Shimizu C, Hojo T, Akashi-Tanaka S, Kinoshita T, Yonemori K, Kouno T, Katsumata N, Ando M, Aogi K, Koizumi F, Nishio K, Fujiwara Y (2011) FcγR2A and 3A polymorphisms predict clinical outcome of trastuzumab in both neoadjuvant and metastatic settings in patients with HER2-positive breast cancer. *Ann Oncol* **22**: 1302–1307.
- Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**: 5116–5121.
- 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 (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530–536.
- van der Hage JA, van de Velde CJ, Julien JP, Tubiana-Hulin M, Vandervelden C, Duchateau L (2001) Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. *J Clin Oncol* **19**: 4224–4237.
- Wakasugi K, Schimmel P (1999) Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science* **284**: 147–151.
- Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, Mcshane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* **131**: 18–43.
- Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B (2001) Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr* **96**: 96–102.
- Yao ZX, Lu LJ, Wang RJ, Jin LB, Liu SC, Li HY, Ren GS, Wu KN, Wang DL, Kong LQ (2014) Discordance and clinical significance of ER, PR, and HER2 status between primary breast cancer and synchronous axillary lymph node metastasis. *Med Oncol* **31**: 798.



This work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)