

# Comparison of prognostic values between combined immunohistochemical score of estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2, Ki-67 and the corresponding gene expression score in breast cancer

Mutsuko Yamamoto-Ibusuki<sup>1</sup>, Yutaka Yamamoto<sup>1</sup>, Satoko Yamamoto<sup>1</sup>, Saori Fujiwara<sup>1</sup>, Peifen Fu<sup>1,2</sup>, Yumi Honda<sup>3</sup>, Ken-ichi Iyama<sup>3</sup> and Hirotaka Iwase<sup>1</sup>

<sup>1</sup>Department of Breast and Endocrine Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; <sup>2</sup>Department of General Surgery First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, PRC and <sup>3</sup>Department of Surgical Pathology, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan

In the clinical diagnosis of breast cancer, immunohistochemistry panels with estrogen receptor (ER), progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2) and Ki-67 are routinely used, and they have been proposed for the classification of breast tumors into distinct subtypes. Gene expression analysis with formalin-fixed paraffin-embedded material have also become widely available recently, but the prognostic values of corresponding gene panels compared with these four immunohistochemical panels had never tested. We independently evaluated the 5-year relapse risk-estimation scores using semiquantitative data of four immunohistochemical panels (Ku-IHC4 score) and compared these with the results of four-gene expression profiling of formalin-fixed paraffin-embedded specimens (Ku-FFPE4 score) in a consecutive series of 235 primary invasive breast cancer patients. Ku-IHC4 score was revealed to be an independent predictor of recurrence other than Ku-FFPE4 score in a multivariate model analyzed by classical clinical parameters (Ku-IHC4 score vs Ku-FFPE4 score;  $\chi^2$ : 14.2 vs 2.5,  $P$ : 0.0002 vs 0.11). When patients were trichotomized into high-, intermediate- and low-risk groups using the thresholds determined from the approximately calculated 5-year relapse rate, Kaplan–Meier analyses showed a significant difference among the three groups in Ku-IHC4 score (log-rank,  $P < 0.0001$ ), but not in Ku-FFPE4 score. The high-risk group according to Ku-FFPE4 score showed contradictory low recurrence rates (Ku-IHC4 score vs Ku-FFPE4 score, 53.1 vs 24.8%), which might be caused by risk-dependently extended error ranges. We show that the Ku-IHC4 score, consisted with semiquantitative measures of immunohistochemistry, provides better prognostic information than the corresponding quantitative RNA measurements. Prognostication tools such as the Ku-IHC4 score may be potentially useful in screening which patients had better be assessed by further testing using other genes rather than ER, PgR, HER2 and Ki-67 to determine critical aspects of therapeutic decision making.

*Modern Pathology* (2013) 26, 79–86; doi:10.1038/modpathol.2012.151; published online 24 August 2012

**Keywords:** breast cancer; gene expression; risk score

Breast cancer is a heterogeneous and complex disease, characterized by molecular and genetic

diversity, which has resulted in the recognition of several fundamentally different subtypes.<sup>1,2</sup> In clinical practice, the use of immunohistochemistry panels has been proposed for the classification of breast tumors into distinct subtypes as identified by gene expression profiling studies. These panels are described as four protein panels and primarily use antibodies against estrogen receptor (ER), progesterone receptor (PgR), human epidermal

Correspondence: Professor H Iwase, MD, PhD, Department of Breast and Endocrine Surgery, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Kumamoto, Kumamoto 860-8556, Japan.

E-mail: hiwase@kumamoto-u.ac.jp

Received 19 March 2012; revised 22 May 2012; accepted 22 May 2012; published online 24 August 2012

growth factor receptor 2 (HER2) and Ki-67.<sup>3–5</sup> Cheang *et al*<sup>6</sup> proposed a simplified classification in which the subtypes defined by clinicopathological criteria are supposed to be similar, but not identical, to intrinsic subtypes, and which represents a convenient approximation. Cuzick *et al*<sup>7</sup> have recently presented a scoring system based on these four assays using detailed semi-quantitative parameters: this would be known as the 'IHC4 score', and is calculated by a proportional hazards regression using classical clinical variables and the four immunohistochemistry values using Trans ATAC datasets for prognostication of ER-positive patients. Moreover, this IHC4 score showed statistically similar performance to that of Oncotype DX (Genomic Health, Redwood City, CA, USA),<sup>8</sup> the most widely used and validated gene-expression assay. The Oncotype DX employs real-time quantitative PCR (RT-qPCR) on formalin-fixed paraffin-embedded specimens, and gene-expression assays such as these are now being emphasized as an important tool for clinical decision making and are recommended by major guidelines.<sup>9,10</sup> Additionally, recent reports have increasingly revealed details about this gene expression profiling tool, for example, that HER2 assessment by Oncotype DX leads to a higher number of cases considered as equivocal,<sup>11</sup> whereas its excellent concordance between ER, PgR and HER2 assessments and immunohistochemical analysis has been reported.<sup>12,13</sup> On the other hand, Oncotype DX also includes analysis of gene expression levels of the genes used in the 'IHC4 score'; nevertheless its head-to-head comparison with four immunohistochemical panels is not clear.

Recently, we independently developed a gene expression analysis system based on material and by using RT-qPCR assays that showed excellent concordance between ER and PgR assessment.<sup>14</sup> Gene expression analysis by RT-qPCR always allows convenient quantification of target transcripts, resulting in data generated as objective continuous variables and more particular reproducibility than immunohistochemistry assessed by manual counting, which is one of its great potential advantages as a method of clinical examination. In this study, we integrated our own scoring models using four panel immunohistochemistry (Ku-IHC4 score) and the corresponding expression levels of the four genes in formalin-fixed paraffin-embedded specimens (Ku-FFPE4 score), with the results of individual multivariate modeling, and compared their prognostic values.

## Patients and methods

### Breast Cancer Tissues

Breast tumor specimens from 235 therapy-naive female patients with primary breast cancer, who

were treated at the Kumamoto University Hospital between 2000 and 2008, were included in this study. The median age of the patients was 59 years (range, 27–93). All patients had undergone surgical treatment. The ethics committee of the Kumamoto University Graduate School of Medical Sciences approved this study. The study was reported according to the Reporting Recommendations for Tumor Marker Prognostic Studies criteria.<sup>15</sup> Neoadjuvant treatments were administered to 42 patients (29 for chemotherapy and 13 for hormonal therapy), and the choice of treatment was decided by risk evaluation according to tumor biology (ER, PgR and HER2 except Ki-67) and clinical staging, including preoperative sentinel lymph node biopsy.<sup>16</sup> Patients received either breast-conservation surgery or total mastectomy and sentinel lymph node biopsy or axillary lymph node dissection. Neoadjuvant and adjuvant treatment and radiotherapy were performed in accordance with the recommendations of the St Gallen international expert consensus on the primary therapy of early breast cancer,<sup>17–20</sup> and included 77.9% treated with hormonal therapy, 37.4% with chemotherapy and 6.4% with targeted therapy using trastuzumab. In addition, breast radiation was administered to 91.0% of the patients (details are described in Table 1). Clinical follow-up included history-taking, physical examinations, laboratory tests and radiologic imaging every 3–12 months for detection of relapse. The median follow-up period was 44 months (range, 4–90).

### Immunohistochemical Analysis for the Ku-IHC4 Score and Gene Expression Analysis for the Ku-FFPE4 Score

Histological sections (4  $\mu$ m) were deparaffinized and incubated for 10 min in methanol containing 0.3% hydrogen peroxide. Sections were stained with rabbit monoclonal antibodies against ER $\alpha$  (SP1, Ventana Japan, Tokyo, Japan) and PgR (1E2, Ventana Japan), HER2 (4B5, Ventana Japan) and Ki-67 (MIB-1, Ventana Japan), and all immunohistochemical procedures were carried out in the NexES IHC Immunostainer (Ventana Medical Systems, Tucson, AZ, USA), in accordance with the manufacturer's instructions (Ki-67 immunostaining was performed retrospectively after the decision regarding therapy had been taken). ER and PgR status was evaluated by percentage of nuclear staining (0–100%), and considered positive when there was  $\geq 1\%$  nuclear staining. HER2 immunostaining was evaluated using the same method as the HercepTest (Dako Japan, Tokyo, Japan); the membranous staining was scored on a scale of 0 to 3+. Accurate number of IHC judgment was as follows: 0 for 10 patients (4%), 1+ for 118 patients (50%), 2+ for 80 patients (23%) and 3+ for 27 patients (11%). Tumors with scores of  $\geq 3$  or 2+ with a  $\geq 2.2$ -fold increase in HER2 gene amplification as determined by fluorescence *in situ* hybridization (FISH) were considered to be positive

**Table 1** Clinicopathological factors

Clinical parameters	Number of patients (%) or median (25%, 75%)
Observation time (months)	44 (0.36, 112.2)
Age (years)	60 (50.0, 70.0)
<i>Menopause</i>	
Pre	60 (26)
Post	175 (74)
<i>Tumor size (mm)</i>	
≤ 20	114 (29)
> 20	121 (51)
<i>Nodal status</i>	
–	147 (63)
+	88 (37)
<i>Clinical stage</i>	
I	93 (40)
II	121 (51)
III	21 (9)
<i>Histological grade</i>	
1	124 (53)
2	50 (21)
3	60 (26)
<i>Lymphovascular invasion</i>	
–	160 (68)
+	59 (25)
<i>ERα</i>	
–	48 (20)
+	187 (80)
<i>PgR</i>	
–	72 (31)
+	163 (69)
<i>HER2</i>	
–	205 (87)
+	30 (13)
Ki-67 Labeling index	24.5 (12.5, 41.1)
<i>Tumor subtype<sup>a</sup></i>	
HR(+), HER2(–)	178 (75)
HR(+), HER2(+)	11 (5)
HR(–), HER2(+)	18 (8)
HR(–), HER2(–)	28 (12)
<i>Endocrine therapy</i>	
No	52 (22)
AI	119 (51)
TAM	17 (7)
TAM-AI	12 (5)
OFS + TAM	35 (15)
<i>Chemotherapy</i>	
No	147 (63)
ACR + T	47 (20)
ACR	28 (12)
T	8 (3)
Other	5 (2)
<i>Trastuzumab treatment</i>	
No	220 (94)
Yes	15 (6)

**Table 1** (Continued)

Clinical parameters	Number of patients (%) or median (25%, 75%)
<i>Breast operation</i>	
Partial mastectomy	173 (74)
Total mastectomy	62 (26)
<i>Radiotherapy</i>	
No	78 (33)
Yes	157 (67)

Abbreviations: ACR, anthracycline-contained regimens; AI, aromatase inhibitors; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; OFS, ovarian function suppression; PgR, progesterone receptor; TAM, tamoxifen; T, taxanes.

<sup>a</sup>HR(+): estrogen receptor (+) and/or progesterone receptor (+).

for HER2 overexpression. Recount was adapted to FISH equivocal cases (from  $\geq 1.8$  to  $< 2.2$ -fold) to obtain a final judgment of HER2 status. Patients of IHC 2+ with FISH positive were only 2 out of 80 (3%). Overall, HER2-positive patients were 29 (12%). Ki-67 was scored for the percentage of cells with nuclear staining cells out of  $1 \times 10^3$  cancer cells in the invasive front of the tumor at  $\times 40$  high-power magnification (Ki-67 labeling index). Details of our gene expression analysis system based on formalin-fixed paraffin-embedded material analyzed by RT-qPCR have been described previously.<sup>14,21</sup> All the primers and probes were purchased from Applied Biosystems (Applied Biosystems Japan, Tokyo, Japan; Hs01046815\_m1 for *ESR1*, Hs01556702\_m1 for *PGR*, Hs01001580\_m1 for *ERBB2*, Hs01032443\_m1 for *MKI67*, Hs01060665\_g1 for *ACTB*, Hs00982775\_m1 for *PUM1*, Hs00359540\_m1 for *TAF10* and Hs00910471\_m1 for *FKBP15*).

### Computation of Risk Scores and Statistics

We generated a model for the estimation of recurrence using gene expression data from our entire cohort. The Cox proportional regression coefficients were computed for ER, PgR, HER2 (assessed by immunohistochemistry/FISH) and Ki-67, and *ESR1*, *PGR*, *ERBB2*, *MKI67* using continuous data. The contribution of each of four variables and these scores were also evaluated by likelihood ratio  $\chi^2$  using the Cox proportional regression model along with univariate and multivariate analyses of prognostic values. A risk score derived from protein assay and gene expression results were then calculated for each patient by calculating the total of the powered products of relative risk by unit for each parameter. Patients were trichotomized into high-, intermediate- and low-risk groups using the threshold determined from their approximately calculated 5-year relapse rate of over 50%, 10–50%, and 0–10%, respectively. Relapse-free survival curves were generated using the Kaplan–Meier method and verified by the log-rank (Mantel–Cox) test. JMP software version 8.0.1 for

**Table 2** Univariate and multivariate analysis for relapse-free survival using each of four variables (Cox proportional regression model)

	Univariate			Multivariate		
	$\chi^2$	P-value	Relative risk (95% CI)	$\chi^2$	P-value	Relative risk (95% CI)
<i>IHC</i>						
ER $\alpha$	28.5	<0.0001	0.97 (0.96, 0.98)	8.61	0.008	0.97 (0.95, 0.99)
PgR	10.4	0.0012	0.98 (0.96, 0.99)	0.012	0.82	1.00 (0.98, 1.02)
HER2	2.02	0.13	2.32 (0.75, 6.00)	1.50	0.22	0.53 (0.16, 1.45)
Ki-67	20.2	<0.0001	38.7 (8.19, 190.26)	2.11	0.15	3.98 (0.62, 30.3)
<i>FFPE</i>						
ESR1	7.34	0.0069	0.90 (0.77, 0.99)	3.27	0.072	0.95 (0.84, 1.00)
PGR	9.25	0.0022	0.92 (0.98, 1.08)	3.79	0.050	0.95 (0.87, 1.00)
ERBB2	5.85	0.015	1.00 (1.01, 0.99)	4.22	0.040	1.00 (1.0001, 1.005)
MKI67	11.2	0.0008	1.01 (1.02, 0.98)	7.54	0.006	1.01 (1.004, 1.02)

Abbreviations: CI, confidential interval; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; PgR, progesterone receptor.

All the parameters were continuous variables, with the relative risk for relapse-free survival calculated per units of each score.

Windows (SAS Institute Japan, Tokyo, Japan) was used for all statistical analysis.

## Results

The clinical characteristics of the 235 cases analyzed in this study are summarized in Table 1. In the analysis of relapse-free survival, local recurrences and distant metastases were considered as events. Of the 21 (9%) recurrent cases among the 235 patients who were available with their complete protein and gene expression data, there were 18 cases (86%) of distant metastases and 3 (14%) cases of local recurrence. Fourteen (67%) of the 21 recurrent patients had died as a result of breast cancer.

### Contribution of Individual Markers

We performed univariate and multivariate analysis for relapse-free survival using a Cox proportional regression model (Table 2). Among the immunohistochemistry variables, ER showed higher prognostic value both in univariate analysis ( $\chi^2=28.5$ ,  $P<0.0001$ ) and multivariate analysis ( $\chi^2=8.61$ ,  $P=0.008$ ), whereas both Ki-67 and PgR, which were significant by univariate analysis ( $\chi^2=20.2$ ,  $P<0.0001$  for Ki-67,  $\chi^2=10.4$ ,  $P<0.0012$  for PgR), had lost their significance in multivariate analysis ( $\chi^2=2.11$ ,  $P=0.15$  for Ki-67,  $\chi^2=0.012$ ,  $P=0.82$  for PgR). On the other hand, *MKI67* showed small but higher prognostic values ( $\chi^2=11.2$ ,  $P=0.0008$  in univariate analysis;  $\chi^2=7.54$ ,  $P=0.006$  in multivariate analysis) among the formalin-fixed paraffin-embedded variables, whereas *ESR1* and *PGR* showed almost equal values ( $\chi^2=7.34$ ,  $P=0.0069$  for *ESR1*,  $\chi^2=9.25$ ,  $P=0.0022$  for *PGR* in univariate analysis;  $\chi^2=3.27$ ,  $P=0.072$  for *ESR1*,  $\chi^2=3.79$ ,  $P=0.050$  for *PGR* in multivariate analysis). HER2 status diagnosed by immunohistochemistry/FISH showed reduced prognostic power in both univariate ( $\chi^2=2.02$ ,  $P=0.13$ ) and multivariate ( $\chi^2=1.50$ ,  $P=0.22$ )

analysis for immunohistochemistry compared with its gene expression data of from formalin-fixed paraffin-embedded specimens ( $\chi^2=5.85$ ,  $P=0.015$  in univariate analysis;  $\chi^2=4.22$ ,  $P=0.040$  in multivariate analysis).

### Creation of Ku-IHC4 and Ku-FFPE4 Scores

Next we developed and defined the Ku-IHC4 score and Ku-FFPE4 score to achieve consistency with each of the four variables by calculating the summation of the powered products of multivariate relative risk by unit for each parameter.

$$\text{KuIHC4 score} = 0.9728^{\text{ER}} + 1.0012^{\text{PgR}} + 0.5269^{\text{HER2}} + 3.9844^{\text{Ki67}}$$

$$\text{KuFFPE4 score} = 0.9525^{\text{ESR1}} + 0.9497^{\text{PGR}} + 1.0029^{\text{ERBB2}} + 1.0116^{\text{MKI67}}$$

The distributions of patient scores are shown as the gray shadowgram in the background of Figure 1. The Ku-IHC4 score ranges from 2.74 to 6.67; median of 3.57 (Figure 1a), and was more normally distributed than the Ku-FFPE4 score (range from 2.25 to 78.5; median of 3.78 (Figure 1b)). The correlation coefficient between these two scores was 0.49 (Spearman rank correlation,  $P<0.0001$ ; data not shown). The approximate likelihood of recurrence at 5 years increased continuously as each score increased. The approximate 5-year relapse rates of 10% and 50% gave values of 3.8 and 4.4, respectively, for the Ku-IHC4 score (Figure 1a), and 4.4 and 6.1, respectively, for the Ku-FFPE4 score (Figure 1b). Two-tailed confidence intervals for the likelihood of recurrence were generally smaller ( $\pm 5$ –10%; Figure 1b) for Ku-IHC4 scores and broader ( $\pm$  more than 20%; Figure 1b) for Ku-FFPE4 scores, which seems to reflect their distribution characteristics.

### Comparison Between Ku-IHC4 and Ku-FFPE4 Scores

To assess the contribution of different clinicopathological variables to the prediction of recurrence, the

relationship between Ku-IHC4 score, Ku-FFPE4 score and relapse-free survival was analyzed by Cox proportional regression models for age, menopausal status, tumor size, nodal status, nuclear grade and lymphovascular invasion (Table 3). In univariate analysis, the  $\chi^2$  value of the Ku-IHC4 score was about 4.5 times greater than the Ku-FFPE4 score (Ku-IHC4 score vs Ku-FFPE4 score; 28.0 vs 6.11), which was identified as an independent predictor of recurrence (Ku-IHC4 score vs Ku-FFPE4 score; multivariate  $\chi^2$ : 14.2 vs 2.5,  $P$ : 0.0002 vs 0.11) by multivariate analyses.

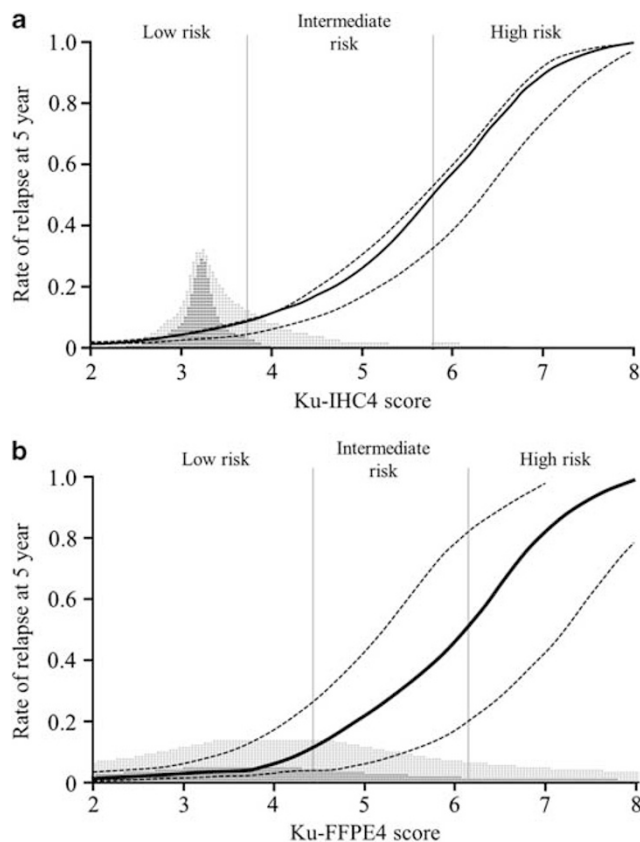
Patients were trichotomized into high-, intermediate- and low-risk groups using the thresholds determined by approximately calculated 5-year relapse rates of over 50% ( $>5.8$  for Ku-IHC4 score,  $>6.1$  for Ku-FFPE4 score: high risk), 10–50% ( $<3.8$  and  $\leq 5.8$  for Ku-IHC4 score,  $<4.4$  and  $\leq 6.1$  for Ku-FFPE4 score: intermediate risk) and 0–10% ( $\leq 3.8$  for Ku-IHC4 score,  $\leq 4.4$  for Ku-FFPE4 score: low risk; Figure 1a and b).

Kaplan–Meier analyses showed a significant difference in recurrence rates between low-, intermediate- and high-risk groups according to Ku-IHC4 score (log-rank correlation  $P < 0.0001$ ; Figure 2a). In contrast, there were no risk-dependent differences in Ku-FFPE4 score especially between intermediate- and high-risk groups (log-rank correlation  $P = 0.78$ ; Figure 2b). The approximately calculated 5-year relapse rates were higher in the low- and intermediate- Ku-FFPE4 score groups (Ku-IHC4 score vs Ku-FFPE4 score: 1.93 vs 6.13% for low-risk group, Ku-IHC4 score vs Ku-FFPE4 score, 22.1 vs 31.53% for intermediate-risk group) than Ku-IHC4 score, but this finding was inverted in high-risk groups due to the contradictory low recurrence rate of the Ku-FFPE4 score (Ku-IHC4 score vs Ku-FFPE4 score, 53.1 vs 24.8%; Supplementary Table S1).

## Discussion

Although the use of multigene profiles has attracted the most attention over recent years for the prediction of clinical outcome, the importance of standardized histopathological analysis of tumors has also been emphasized. In our study, we compared the prognostic information provided by two methods of prognostic scoring composed of ER, PgR, HER2 and Ki-67 analysis by immunohistochemistry, and their corresponding quantitative RNA measurements, and found that semiquantitative measures of four immunohistochemistry panels provide higher prognostic information.

In multivariate analysis of the components of the Ku-IHC4 score, the prognostic value of ER was so great that the other three components were counterbalanced; in contrast the Ku-FFPE4 score was composed of uniform prognostic values. We previously reported a high concordance rate of 97% between *ESR1* levels in FFPE samples and ER



**Figure 1** Estimated probability of relapse as a continuous function of the Ku-IHC4 score (a) and the Ku-FFPE4 score (b). The continuous relationship between the respective score and the probability of developing a recurrence within the first 5 years after starting systematic therapies for breast cancer is described by an independent model for each score. The curves indicate 95% CI. The gray shadowgram in the background shows the distribution of scores for the patients.

expression in immunohistochemistry, and moderate concordance between *PGR* gene expression and PgR immunohistochemical expression with a concordance rate of 83%.<sup>14</sup> These small inconsistencies may be one of the causes of the reduced prognostic value of the Ku-FFPE4 score, due to factors such as variation in the proportion of tumor cells, as well as variation in the specific cells around the tumor such as lymphocytes and mesenchymal cells that may persist in spite of careful macrodissection. The data derived from immunohistochemistry analysis were the result of specific evaluation of the tumor cells. If the tumor characteristics mostly depend not on the transcript levels but on the protein expression of certain molecules, immunohistochemistry is more suitable for prognostication. With regard to Ki-67, our previous data demonstrated that the *MKI67* gene expression of formalin-fixed paraffin-embedded specimens showed similar patterns of significance in Kaplan–Meier analysis and univariate and multivariate RFS results to the Ki-67 labeling index, but were not superior to the Ki-67 immunohistochemical results.<sup>22</sup> With regard to the concordance

**Table 3** Univariate and multivariate analysis between clinicopathological characteristics, Ku-IHC4 score and Ku-FFPE4 score for relapse-free survival (Cox proportional regression model)

Variables	Univariate			Multivariate with Ku-IHC4 score			Multivariate with Ku-FFPE4 score		
	$\chi^2$	Relative risk (95% CI)	P-value	$\chi^2$	Relative risk (95% CI)	P-value	$\chi^2$	Relative risk (95% CI)	P
Age (</ $\geq$ 50)	2.8	0.5	0.095						
Menopause	2.9	2.2	0.088						
Tumor size (</>20 mm)	8.3	4.2	0.0039	3.8	3.2	0.051	4.6	3.5	0.031
Nodal status	10.4	4.3	0.0013	3.2	2.8	0.076	3.4	2.9	0.067
Nuclear grade	12.2	2.5 (1.5–4.5)	0.019	0.5	1.3 (0.6–2.6)	0.501	5.9	4.4 (1.3–16.3)	0.014
Lymphovascular invasion	4.5	2.6	0.033	0.1	1.2	0.73	$5.9 \times 10^{-4}$	1.0	0.98
Ku-IHC4 score	28.0	2.9 (2.0–4.4)	<0.001	14.2	2.7 (1.6,4.4)	0.0002			
Ku-FFPE4 score	6.1	1.1 (1.0–1.1)	0.013				2.5	16.5 (0.3–218.4)	0.11

Abbreviation: CI, confidential interval.

Age at starting systematic therapy was a binary variable (0 for an age of less than 50 years and 1 for an age of 50 years or more); menopausal status was a binary variable (0 for menopausal and 1 for premenopausal); tumor size was measured by invasive diameter in histopathological specimens and treated as a binary variable (0 for a diameter of 20 mm or less and 1 for a diameter greater than 20 mm); nodal status was a binary variable (0 for negative and 1 for positive); nuclear grade was a continuous variable, lymphovascular invasion was a binary variable (0 for negative and 1 for positive); and the Ku-IHC4 score and the Ku-FFPE4 score were continuous variables, with the relative risk for relapse-free survival calculated per units of each score.

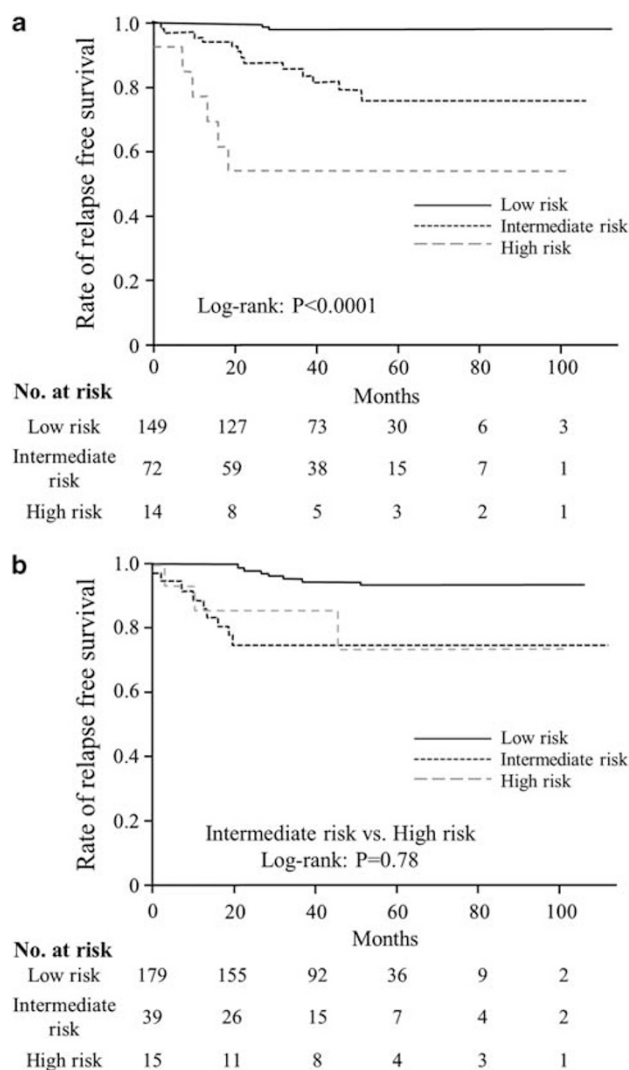
between HER2 immunohistochemistry/FISH status and *ERBB2* gene expression status, our data apparently showed a good correlation between HER2 immunohistochemistry/FISH score and *ERBB2* gene expression levels (Supplementary Figure S1a and b). However, when analyzed in detail, gene expression data ranged from 0.3 to 70.7 in HER2 immunohistochemistry/FISH-negative patients and from 0.7 to 1073 in HER2 immunohistochemistry/FISH-positive patients (Supplementary Figure S1b). Consequently the overlapping zone of gene expression between HER2 immunohistochemistry/FISH-positive and -negative patients was too broad to provide any meaningful results (0.7–70.7,  $n=206$ ). We found a moderate concordance rate of 82.1% between *ERBB2* gene expression and HER2 immunohistochemistry/FISH positivity judged by a cut-off value simply defined from the receiver observer curve (AUC: 0.79). Dobb *et al*<sup>11</sup> reported discordance between HER2 mRNA assessed by Oncotype DX and HER2 gene amplification, due to a high false-negative rate. Although our concordance rate was reasonable compared with other reports concerning the correlation between HER2 immunohistochemistry/FISH and *ERBB2* gene expression,<sup>13,23</sup> more detailed analysis and comparison of our data is required. In addition, it should be borne in mind that differences in immunohistochemical values can naturally occur as a result of variability in several factors including fixation, antigen retrieval, reagents and interpretation.

As discussed above, the small differences between ER, PgR, HER2 and Ki-67 immunohistochemistry and their corresponding gene expression is reflected in the distribution of the Ku-FFPE4 score and explains why these Ku-FFPE4 score showed the risk-dependent broader error ranges (Figure 1b, Supplementary Table S1). Consequently the larger

number of patients in the low-risk group ( $n=179$ , 76.1%; Supplementary Table S1) according to Ku-FFPE4 score, and the smaller number in the intermediate-risk group ( $n=39$ , 16.6%) compared with the Ku-IHC4 score may result in failure to discriminate especially between intermediate- and high-risk groups, when using the Ku-FFPE4 score to assess prognosis (Figure 2b).

More interestingly, HER2 status was defined as a risk reduction factor (relative risk 0.53; Table 2) according to the Ku-IHC4 score, but not the Ku-FFPE4 score (relative risk 1.00; Table 2). We speculate that regular use of trastuzumab for HER2-positive patients in a neoadjuvant and adjuvant setting may have caused their prognoses to be dramatically inverted in the past decade. However, we should remember that prognostication by Ku-IHC4 score might be influenced through our standard therapies on the basis of immunohistochemical methods (especially in ER, PgR and HER 2), which might correlate with the superior prognostic values of the Ku-IHC4 score. In addition immunohistochemistry assessments of Ki-67, a simple proliferation surrogate, were retrospectively added following St Gallen's recommendation in 2009,<sup>4</sup> and protein expression of Ki-67 rather than its gene expression has proven to be a useful prognostic marker. Overall, our Ku-IHC4 score is defined as the risk-estimation tool for patients who have undergone standard therapies, thus intermediate-risk and high-risk groups may need to undergo additional testing to aid therapy selection, such as use of a multigene assay covering genes other than these four genes, or use of other new therapeutic agents currently in development.

An advantage of the Ku-IHC4 score is that evaluation of all the serial parameters consisting of ER, PgR, HER2 and Ki-67 are conducted in the routine workup of our hospital and judged by



**Figure 2** Likelihood of relapse according to each Ku-IHC4 score (a) and Ku-FFPE4 score (b) category. A low risk was defined as an estimated 5-year relapse rate of <10%, an intermediate risk as an estimated 5-year relapse rate of between 10 and 50%, and a high risk as an estimated 5-year relapse rate of over 50%. The difference among the groups was significant ( $P < 0.001$ ) in Ku-IHC4 score, in contrast the intermediate and high-risk groups in the Ku-FFPE4 score were not significantly different ( $P = 0.78$ ).

surgical pathologists who are also in charge of other general works. The Ku-IHC4 score is also attractive because it is far less expensive than the multigene assay and uses assays performed as a part of the current standard of care. Before widespread implementation of this kind type of risk-estimation tool, however, standard guidelines for immunohistochemistry have to be agreed, such as the American Society of Clinical Oncologists/College of Pathologists guidelines for ER, PgR and HER2 testing,<sup>24,25</sup> while accurate guidelines are still required for Ki-67.<sup>26</sup> Moreover, multivariate analysis with Ku-IHC4 score showed a prognostic value relatively independent of tumor size ( $\chi^2 = 3.78$ ,  $P = 0.051$ ; Table 3) and nodal status ( $\chi^2 = 3.15$ ,  $P = 0.076$ ;

Table 3), and thus has great potential to create a prognostic model that integrates this information with classical clinical and pathological variables that have proven to be helpful. Further, an examination of the impact of interlaboratory variability in immunohistochemical staining will be required, and, thus, would benefit from evaluation of the immunohistochemistry four scores in several independent sample/data sets.

In conclusion, although it is well known that multiple gene expression assays have provided additive information for breast cancer patients, the best utilization of existing classical variables are reconfirmed to be important. Prognostication tools such as the Ku-IHC4 score may be potentially useful in screening which patients had better be assessed by further testing using other genes rather than ER, PgR, HER2 and Ki-67 to determine critical aspects of therapeutic decision making.

## Acknowledgements

We thank Y Azakami for excellent technical support and A Okabe for excellent clinical data management.

## Disclosure/conflict of interest

The authors declare no conflict of interest.

## References

- 1 Perou CM, Sorlie T, Eisen MB, *et al*. Molecular portraits of human breast tumours. *Nature* 2000;406:747–752.
- 2 van 't Veer LJ, Dai H, van de Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536.
- 3 Carey LA, Perou CM, Livasy CA, *et al*. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492–2502.
- 4 Goldhirsch A, Ingle JN, Gelber RD, *et al*. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol* 2009; 20:1319–1329.
- 5 Goldhirsch A, Wood WC, Coates AS, *et al*. 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* 2011;22:1736–1747.
- 6 Cheang MC, Chia SK, Voduc D, *et al*. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* 2009;101:736–750.
- 7 Cuzick J, Dowsett M, Pineda S, *et al*. 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* 2011;29:4273–4278.
- 8 Paik S, Shak S, Tang G, *et al*. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–2826.

- 9 Harris L, Fritsche H, Mennel R, *et al*. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 2007;25:5287–5312.
- 10 Engstrom PF, Bloom MG, Demetri GD, *et al*. NCCN molecular testing white paper: effectiveness, efficiency, and reimbursement. *J Natl Compr Canc Netw* 2011;9(Suppl 6):S1–S16.
- 11 Dabbs DJ, Klein ME, Mohsin SK, *et al*. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the Oncotype DX test: an independent quality assurance study. *J Clin Oncol* 2011;29:4279–4285.
- 12 Badve SS, Baehner FL, Gray RP, *et al*. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol* 2008;26:2473–2481.
- 13 Baehner FL, Achacoso N, Maddala T, *et al*. Human epidermal growth factor receptor 2 assessment in a case-control study: comparison of fluorescence in situ hybridization and quantitative reverse transcription polymerase chain reaction performed by central laboratories. *J Clin Oncol* 2010;28:4300–4306.
- 14 Ibusuki M, Fu P, Yamamoto S, *et al*. Establishment of a standardized gene-expression analysis system using formalin-fixed, paraffin-embedded, breast cancer specimens. *Breast Cancer* 2011 (e-pub ahead of print).
- 15 McShane LM, Altman DG, Sauerbrei W, *et al*. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol* 2005;2:416–422.
- 16 Iwase H, Yamamoto Y, Kawasoe T, *et al*. Advantage of sentinel lymph node biopsy before neoadjuvant chemotherapy in breast cancer treatment. *Surg Today* 2009;39:374–380.
- 17 Goldhirsch A, Glick JH, Gelber RD, *et al*. Meeting highlights: International Consensus Panel on the treatment of primary breast cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol* 2001;19:3817–3827.
- 18 Goldhirsch A, Wood WC, Gelber RD, *et al*. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003;21:3357–3365.
- 19 Goldhirsch A, Glick JH, Gelber RD, *et al*. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005;16:1569–1583.
- 20 Goldhirsch A, Wood WC, Gelber RD, *et al*. Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. *Ann Oncol* 2007;18:1133–1144.
- 21 Fu P, Ibusuki M, Yamamoto Y, *et al*. Quantitative determination of insulin-like growth factor 1 receptor mRNA in formalin-fixed paraffin-embedded tissues of invasive breast cancer. *Breast Cancer* 2011 (e-pub ahead of print).
- 22 Yamamoto S IM, Yamamoto Y, Fu P, *et al*. Clinical relevance of Ki67 gene expression analysis using formalin-fixed and embedded in paraffin (FFPE) breast cancer specimens. *Breast Cancer* 2012 (e-pub ahead of print).
- 23 Barberis M, Pellegrini C, Cannone M, *et al*. Quantitative PCR and HER2 testing in breast cancer: a technical and cost-effectiveness analysis. *Am J Clin Pathol* 2008;129:563–570.
- 24 Hammond ME, Hayes DF, Dowsett M, *et al*. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010;28:2784–2795.
- 25 Middleton LP, Price KM, Puig P, *et al*. Implementation of American Society of Clinical Oncology/College of American Pathologists HER2 Guideline Recommendations in a tertiary care facility increases HER2 immunohistochemistry and fluorescence in situ hybridization concordance and decreases the number of inconclusive cases. *Arch Pathol Lab Med* 2009;133:775–780.
- 26 Dowsett M, Nielsen TO, A'Hern R, *et al*. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst* 2011;103:1656–1664.

Supplementary Information accompanies the paper on Modern Pathology website (<http://www.nature.com/modpathol>)