

Comparing subjective and digital image analysis HER2/neu expression scores with conventional and modified FISH scores in breast cancer

I Skaland,¹ I Øvestad,¹ E A M Janssen,¹ J Klos,¹ K H Kjellefold,¹ T Helliesen,¹ J P A Baak^{1,2,3}

¹ Department of Pathology, Stavanger University Hospital, Stavanger, Norway; ² Department of Pathology, The Gade Institute, University of Bergen, Norway; ³ Department of Pathology, Free University, Amsterdam, The Netherlands

Correspondence to: Professor dr J P A Baak, Department of Pathology, Stavanger University Hospital, Box 8100, 4068 Stavanger, Norway; baja@sus.no

Published Online First
30 March 2007

ABSTRACT

Background: HER2/neu expression and fluorescence in situ hybridisation (FISH) amplification have therapeutic significance.

Aims: To compare subjective HER2/neu expression scores with digital image analysis (DIA) and conventional and modified FISH scores in breast cancer.

Methods: Sixty HercepTest-immunostained breast carcinomas, prospectively scored as consensus 2+ and 3+ (DAKO protocol) by two observers, were analysed with DIA, and conventional (Vysis) and modified FISH scoring protocols.

Results: With consensus scoring, 23 (38%) of the 60 cases were 2+ and 37 (62%) were 3+. Agreement with DIA scores was 100%. With conventional FISH scoring, 4 of the 3+ cases did not show amplification, but all of those negative cases had high HER2/neu copy numbers. With the modified FISH scoring protocol, all HercepTest immunohistochemical 3+ cases were amplified. Of the 2+ cases, 3 were amplified with the modified FISH protocol and 4 with the conventional FISH protocol.

Conclusions: Modified FISH scores were better correlated with HercepTest 3+ consensus and DIA scores than were conventional FISH scores. HER2/neu DIA scoring is a cost-effective supplementary tool in surgical pathology.

Breast cancer is a common disease. The prognosis for lymph node-negative patients is favourable, but 15–30% of them develop distant metastases. Chemotherapy can improve prognosis in certain patient subgroups but has undesirable side effects. This has led to intensified efforts for more specific therapeutic targets and therapies. One of the best known therapies is antibody-based therapy against HER-2/neu receptors. Over-expression of HER-2/neu receptors is associated with aggressive biological behaviour and poor clinical outcome. Antibody-based therapy with trastuzumab (Herceptin), can be effective in metastatic breast cancer when taken alone or in combination with chemotherapeutic agents.^{1,2}

Standardisation of an optimal method to evaluate HER-2 expression is necessary to define its role as a therapeutic target in breast cancer. Unfortunately, subjective scoring is not perfectly reproducible across observers.^{3,4} Disagreement in HER-2 expression scoring implies that some patients who are in need of trastuzumab treatment will not receive it, and patients who are insensitive to the effects of trastuzumab will be treated with it.^{2,5} This is unacceptable from the viewpoints of both human suffering and healthcare costs. Therefore, combined detection of HER-2/neu

expression and amplification by immunohistochemistry (IHC) and fluorescent in situ hybridisation (FISH) is used, which are believed to supplement each other in breast cancer HER2/neu classification.

It is generally assumed that 3+ expression patterns are associated with FISH amplification, but in a prospective routine analysis, we discovered that this is not the case. The lack of correlation of Her2 3+ and FISH amplification, may be due to subjective scoring errors, but also shortcomings in the FISH scoring protocol. There are several possible explanations for this, including errors in the staining protocol, inaccurate expression scoring, or errors in the FISH scoring protocol. Therefore, we analysed HER2/neu expression in breast cancer with consensus scoring, highly reproducible digital image analysis (DIA), and conventional and a modified FISH scoring protocol.

MATERIALS AND METHODS

Patients and samples

The study is part of an ongoing breast cancer study in our laboratory, which is approved by the local regional ethics committee, in accordance with the Norwegian Biobank law. HER2/neu is routinely assessed on all invasive breast cancers in our laboratory. In the period May 2005 to July 2006, 60 consecutive breast cancers were consensus-scored by two pathologists as 2+ or 3+ for HER2/neu (Dako protocol) (many more were negative for HER2/neu, but are not considered here). The original tumours were cut into 0.5 cm blocks, fixed in buffered 4% formaldehyde, and embedded in paraffin. Paraffin blocks were sliced at 4 µm and stained with H&E–safran (H&ES). H&ES slides were used for diagnosis and for selecting paraffin blocks for HER2/neu IHC and FISH analyses. Whole tissue sections were analysed.

Immunohistochemistry and subjective expression scoring

A HercepTest kit and Autostainer (Dako, Glostrup, Denmark) were used for IHC analysis, according to the manufacturer's FDA-approved procedures using the American Society of Clinical Oncology 2006 modification (see <http://www.jco.org/cgi/doi/10.1200/jco.2006.09.2775>). Subjective expression scores were based on the Dako protocol as follows; 0, no membrane staining; 1+, faint partial membrane staining; 2+, weak-to-moderate complete membrane staining in >10% of the invasive cancer cells; 3+, intense complete membrane



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jcp.bmj.com/info/unlocked.dtl>

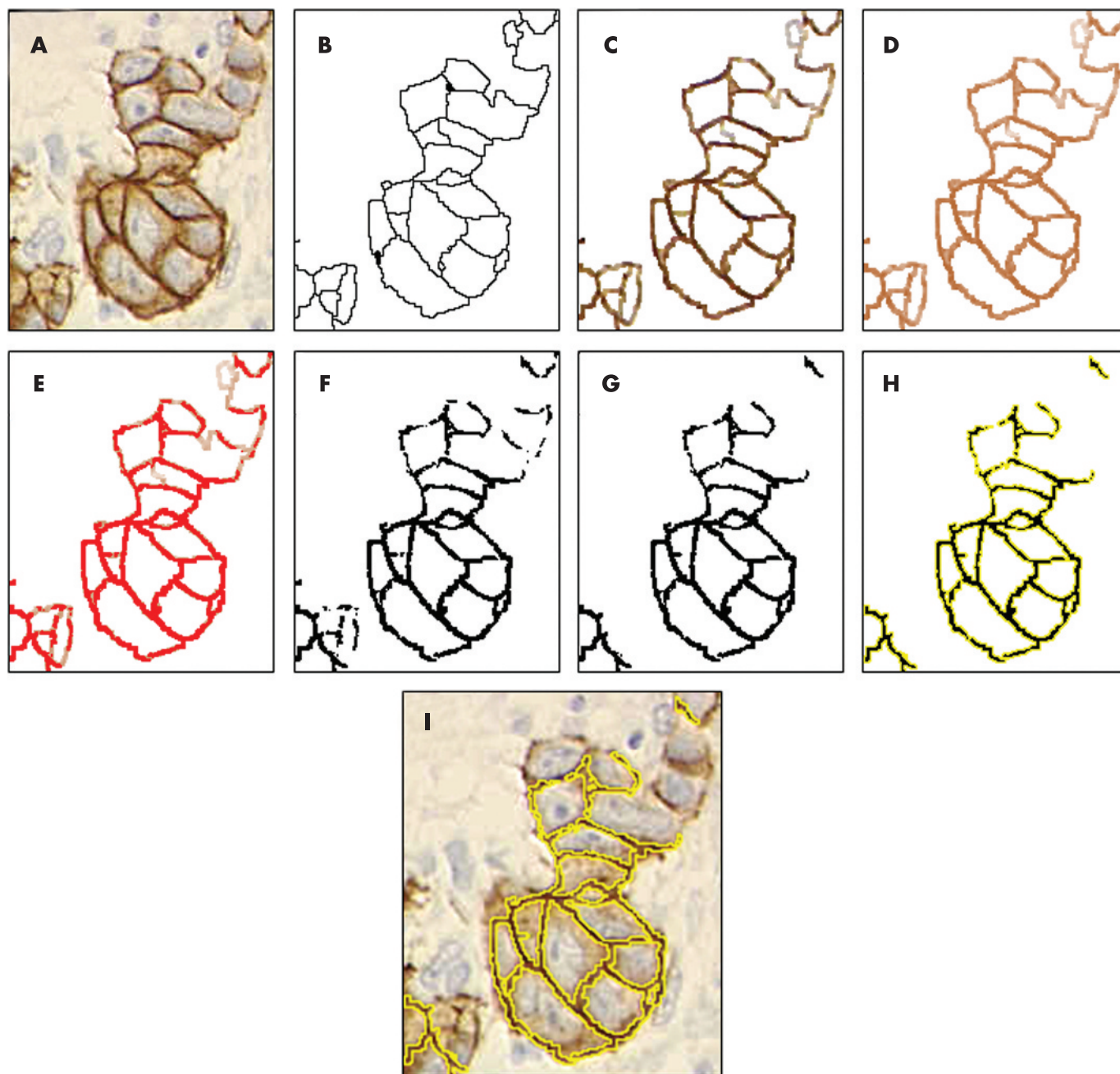


Figure 1 Schematic illustration showing how membranes are segmented by the automated digital image analysis method. (A) Original HER2/neu immunohistochemically stained image. (B) Result of find maxima and converting to mask. (C) Result after a selection of the mask in B is loaded into A and all picture points outside the mask are deleted. (D) brown diaminobenzidine component of the colour deconvoluted image in C. (E) Threshold of D. (F) Mask created based on threshold in E. (G) Same as F but small objects are removed. (H) Selection created on the basis of G. (I) Selection from H loaded into the original image where the measurement is made.

Table 1 Correlation between subjective consensus scores of two pathologists and digital image analysis (DIA) scores (using grey scale = 119 as decision threshold) ($p < 0.001$)

	DIA score		Total
	2+	3+	
Subjective consensus score			
2+	23	0	23
3+	0	37	37
Total	23	37	60

staining in $>10\%$ of the invasive cancer cells. Two pathologists performed the scoring. The two have worked closely for many years and regularly discussed the degree of HER-2 staining. Subjective consensus scores were obtained for each breast cancer using a multiheaded microscope and without knowledge of prior scores and FISH results.

DIA

Slides were evaluated using the ImageJ DIA program (National Institutes of Health, USA; <http://rsb.info.nih.gov/ij/>). For each

Table 2 Correlation between the subjective consensus scores, and conventional Vysis fluorescence in situ hybridisation (FISH) scores ($p < 0.001$)

	FISH score (Vysis scoring protocol)			Total
	Neg	Low amp	Amp	
Subjective consensus score				
2+	19	2	2	23
3+	4	1	9	14
Total	23	3	11	37

Neg, HER2/neu/CEP17 ratio < 1.8 ; Low amp, ratio 1.8–2.2; Amp, ratio > 2.2 .

tumour, from the areas of strongest HER2/neu staining of the invasive front, five fields of vision (FOV) were selected and digitised with a 40 \times objective (NA 0.75) and a Scion Image (CFW-1312C) colour CCD camera using a Leica microscope LM DM with a CTRMIC control unit (Leica, Wetzlar, Germany). To ensure stable image quality, the camera system was supplied by a UPS voltage stabiliser (which guaranteed $230V \pm 1V$), and the microscope aperture and light settings were automatically controlled by Leica DM SDK 4.1/MIC control software. Before each image capturing session, the system was calibrated with an optical-density calibration slide (PRESS-PRO21).

Based on built-in ImageJ program functions, we created a macro to automatically segment membrane staining and measure the area of membrane fragment and the mean and median staining intensity. Figure 1 illustrates the different image processing steps. Briefly, membrane staining was segmented by using the *find maxima* (segmented particles) function followed by the *converting to mask* function. The mask was adjusted with two dilations and erosions. Then, a selection was created which was overlaid (i.e. restored) on the original image. Everything outside the selection was deleted, resulting in an image that was comprised almost exclusively of stained membrane. The image was then cleaned up by a combination of *colour deconvoluting* and *thresholding* the deconvoluted image that contained the diaminobenzidine (DAB) component, using the *Otsu threshold*,⁶ and removing membrane fragments < 50 pixels. This resulted in a new image containing only larger DAB-stained membranes and larger membrane fragments. Based on this image, a new mask and selection was made and restored in the original image from which the final measurements were taken. The final selection in the original image could be viewed directly on the screen and manually inspected along with the measurements (fig 1).

The immunohistochemical expression is calculated as a linear grey scale, in which totally white = 255 and totally black = 0.

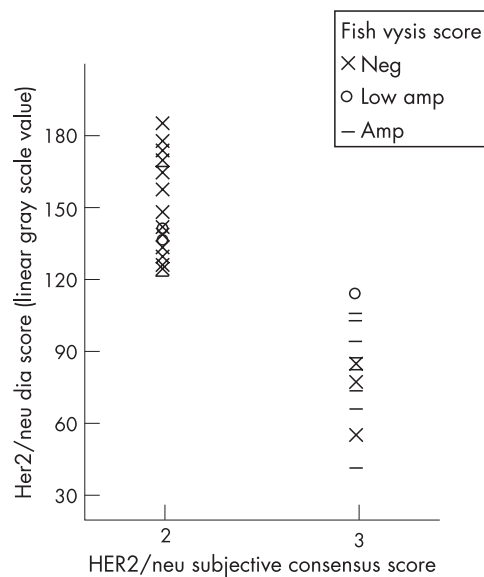
FISH analysis

Since January 2006, FISH analysis has been used routinely, as an add-on to immunohistochemistry, for further HER2-classifica-

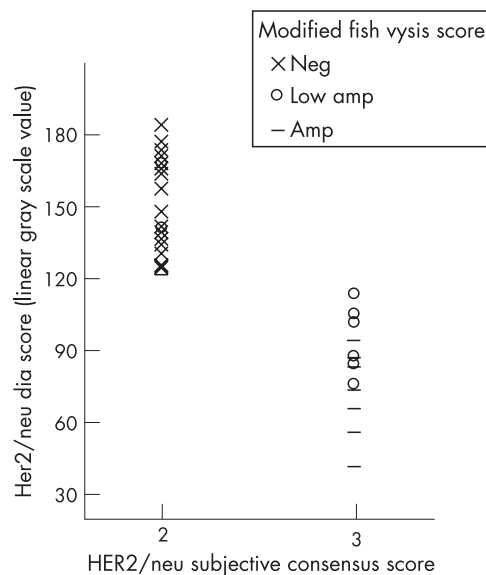
Table 3 Correlation between subjective consensus scores and FISH scores according to the modified protocol ($p < 0.001$)

	FISH score (modified (Zymed) scoring protocol)			Total
	Neg	Low amp	Amp	
Subjective consensus score				
2+	20	2	1	23
3+	0	7	7	14
Total	20	9	8	37

Neg, HER2/neu 0–5 signals/nuclei; Low amp, > 5 –10 signals/nuclei; Amp > 10 signals/nuclei.

**Figure 2** Scatter plot of subjective consensus scores and digital image analysis (DIA) scores, indicating fluorescence in situ hybridisation (FISH) scores according to the Vysis scoring protocol. Neg, HER2/neu/CEP17 ratio < 1.8 ; Low amp, ratio 1.8–2.2; Amp, ratio > 2.2 .

tion of IHC 2+ and 3+ breast cancers. FISH was performed in 37 of the 60 2+/3+ cases. The PathVysion (Vysis Inc, Downers Grove, Illinois, USA) assay was used to evaluate amplification of the HER-2/neu gene, according to the manufacturer's FDA-approved protocol. HER-2/neu and centromere chromosome 17 (CEP17) signals were manually counted using a fluorescent microscope. A HER2/CEP17 ratio of 2.2 or higher was considered positive for HER-2/neu amplification. A HER2/neu/CEP17 ratio < 1.8 was considered negative for HER-2/neu amplification. The few cases with ratios between 1.8 and 2.2 were interpreted as having low amplification. Using this classification rule, repeated independent tests of 25 cases in

**Figure 3** Scatter plot of subjective consensus scores and digital image analysis (DIA) scores, indicating HER2/neu amplification according to the modified fluorescence in situ hybridisation (FISH) scoring protocol. Neg, 0–5 signals/nuclei; Low amp, > 5 –10 signals/nuclei; Amp, > 10 signals/nuclei.

Take-home messages

- ▶ Subjective HER-2 expression scoring in breast cancer has therapeutic significance, but subjective scoring reproducibility is not perfect.
- ▶ Consensus scoring or digital image analysis may improve accuracy and reproducibility of Hercep Test Scoring.
- ▶ Amplification of CEP17 is to be taken into account when scoring dual colour HER-2 FISH.

our laboratory by different technicians and in another laboratory (Ullevål Hospital, Oslo, Norway) showed complete concordance. The FISH scoring was also evaluated with a modified (Zymed chromogenic in situ hybridisation (CISH)) scoring protocol as follows: mean HER2/neu 0–5 signals/nucleus, negative; >5–10 signals/nucleus, low amplification; >10 signals/nucleus, amplification.

Statistical analysis

Correlation tables were made and correlations between variables were calculated using SPSS v.14.0.

RESULTS

Figure 1 outlines the steps for DIA. Note the correlation between actual immunohistochemical staining and DIA-detected HER2/neu expression. Based on consensus scoring, 23 (38%) of the cases were 2+ and 37 (62%) were 3+; these scores were 100% in agreement with DIA scores (table 1). With conventional FISH scoring, 4 (21%) of the 3+ cases showed no amplification (table 2, fig 2), but all of the four negative cases had high HER2/neu copy numbers. With the modified FISH scoring protocol, all HercepTest 3+ cases were FISH positive (table 3, fig 3). Of the 2+ cases, 3 (13%) were amplified according to the modified FISH protocol, and 4 (17%) were amplified according to the conventional protocol (tables 2 and 3). None of the DIA 3+ cases was negative according to the modified FISH protocol.

DISCUSSION

Agreement between the subjective consensus scores and the DIA was excellent, as was the correlation between IHC 3+ and FISH when the modified FISH scoring protocol was used. However, the correlation was not as good (71%) when the Vysis dual-colour protocol was used. The Vysis protocol score is the ratio of the number of HER2/neu and CEP17 signals. In the four IHC 3+ negative Vysis FISH cases, the number of CEP17 signals was higher than the expected 1–2 signals or 3–5 polysomy signals. It is noted that >5 HER2/neu signals correlates with a 2.5c exceeding rate, which in ploidy analysis is a widely accepted high-risk indicator for a variety of diagnostic and prognostic applications.^{7–9} In the four Vysis FISH negative HercepTest 3+ cases studied, the average HER-2 signal amplification range was 6–10 CEP17 and 6–11 HER2/neu signals/nuclei. These are most likely cases that have amplified CEP17 and HER2/neu genes, but the ratio of the CEP17 and HER2/neu scores is <1.8, despite clear HER2/neu amplification.

This would explain the negative Vysis FISH classification and the three low amplification cases and one amplified case with the modified FISH protocol. Amplification of CEP17 is to be taken into account when scoring dual colour FISH, since this is reported to occur in a variable number of breast cancers. It has been recommended¹⁰ that the raw data number of HER2/neu signals and CEP17 signals be evaluated together with the ratio to avoid misinterpretation of rare cases with high CEP17 and HER2/neu copy numbers. Hofmann *et al*¹¹ described three (1.0%) of 302 patients with FISH negative ICH 3+ positive tumours who responded to trastuzumab therapy, and noted that two of the cases had >9 CEP17 signals/nucleus. Therefore, it is likely that the four IHC 3+ cases with high copy numbers of HER2/neu and CEP17 were correctly scored as 3+ by consensus, DIA, and the modified FISH (Zymed CISH) scoring protocol.

Many studies have shown considerable inter- and intra-observer variability in subjective IHC HER2/neu scores.^{3–4} Although consensus scoring appears to increase the accuracy and reproducibility of subjective scores, it is unrealistic in a clinical setting to use labour-intensive consensus scoring with several expert breast pathologists in each IHC 2 and 3+ case, especially in small laboratories. The current results show that HER2/neu scoring by DIA is accurate, time-effective (5–10 min/assessment), reproducible, and does not require expensive equipment (the costs of the system we used was about €6000 (US\$8000, or £4000)).

We conclude that modified FISH scores correlate better than conventional FISH scores with HercepTest 3+ consensus and DIA scores. Furthermore, HER2/neu DIA scoring is a cost-effective supplementary tool in surgical pathology.

Competing interests: None.

REFERENCES

1. **Perez EA.** HER-2 as a prognostic, predictive, and therapeutic target in breast cancer. *Cancer Control* 1999;**6**:233–40.
2. **Gonzalez-Angulo AM, Hortobágyi GN, Esteva FJ.** Adjuvant therapy with trastuzumab for HER-2/neu-positive breast cancer. *Oncologist* 2006;**11**:857–67.
3. **Bloom K, Harrington D.** Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy. *Am J Clin Pathol* 2004;**121**:620–30.
4. **Rhodes A, Borthwick D, Sykes R, et al.** The use of cell line standards to reduce HER-2/neu assay variation in multiple European cancer centers and the potential of automated image analysis to provide for more accurate cut points for predicting clinical response to trastuzumab. *Am J Clin Pathol* 2004;**122**:51–60.
5. **Sidoni A, Ferri I, Cavaliere A, et al.** Detection of HER-2/neu (c-erbB-2) overexpression and amplification in breast carcinomas with ambiguous immunohistochemical results. A further contribution to defining the role of fluorescent in situ hybridization. *Anticancer Res* 2006;**26**:2333–7.
6. **Otsu N.** A threshold method from gray-level histograms. *Trans Syst Man Cybern* 1979;**9**:62–6.
7. **Sun XR, Wang J, Garner D, et al.** Detection of cervical cancer and high grade neoplastic lesions by a combination of liquid-based sampling preparation and DNA measurements using automated image cytometry. *Cell Oncol* 2005;**27**:33–41.
8. **Mattila R, Alanen K, Syrjänen S.** DNA content as a prognostic marker of oral lichen planus with a risk of cancer development. *Anal Quant Cytol Histol* 2004;**26**:278–84.
9. **Hanselaar AG, Bocking A, Gundlach H, et al.** International Consensus Conference on the Fight Against Cervical Cancer, IAC Task Force 8 Summary, Chicago, Illinois, USA. Summary statement on quantitative cytochemistry (DNA and molecular biology): Task Force 8. *Acta Cytol* 2001;**45**:499–501.
10. **Troxel ML, Bangs CD, Lawse HJ, et al.** Evaluation of Her-2/neu status in carcinomas with amplified chromosome 17 centromere locus. *Am J Clin Pathol* 2006;**126**:709–16.
11. **Hofmann MK, Gross C, Beyser K, et al.** Predictive value of HercepTest IHC and Pathvysion FISH data: analysis from a trastuzumab phase II monotherapy study (abstract 740). *Proc ASCO* 2003;**22**:185.