

## COMMENTARY

# Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group

Mitch Dowsett, Torsten O. Nielsen, Roger A'Hern, John Bartlett, R. Charles Coombes, Jack Cuzick, Matthew Ellis, N. Lynn Henry, Judith C. Hugh, Tracy Lively, Lisa McShane, Soon Paik, Frederique Penault-Llorca, Ljudmila Prudkin, Meredith Regan, Janine Salter, Christos Sotiriou, Ian E. Smith, Giuseppe Viale, Jo Anne Zujewski, Daniel F. Hayes

Manuscript received March 14, 2011; revised September 1, 2011; accepted September 2, 2011.

**Correspondence to:** Mitch Dowsett, BSc, PhD, Department of Biochemistry, Royal Marsden Hospital and Breakthrough Breast Cancer Centre, Fulham Rd, London SW3 6JJ, UK (e-mail: [mitch.dowsett@icr.ac.uk](mailto:mitch.dowsett@icr.ac.uk)).

Uncontrolled proliferation is a hallmark of cancer. In breast cancer, immunohistochemical assessment of the proportion of cells staining for the nuclear antigen Ki67 has become the most widely used method for comparing proliferation between tumor samples. Potential uses include prognosis, prediction of relative responsiveness or resistance to chemotherapy or endocrine therapy, estimation of residual risk in patients on standard therapy and as a dynamic biomarker of treatment efficacy in samples taken before, during, and after neoadjuvant therapy, particularly neoadjuvant endocrine therapy. Increasingly, Ki67 is measured in these scenarios for clinical research, including as a primary efficacy endpoint for clinical trials, and sometimes for clinical management. At present, the enormous variation in analytical practice markedly limits the value of Ki67 in each of these contexts. On March 12, 2010, an international panel of investigators with substantial expertise in the assessment of Ki67 and in the development of biomarker guidelines was convened in London by the cochair of the Breast International Group and North American Breast Cancer Group Biomarker Working Party to consider evidence for potential applications. Comprehensive recommendations on preanalytical and analytical assessment, and interpretation and scoring of Ki67 were formulated based on current evidence. These recommendations are geared toward achieving a harmonized methodology, create greater between-laboratory and between-study comparability, and allow earlier valid applications of this marker in clinical practice.

J Natl Cancer Inst 2011;103:1656–1664

Uncontrolled proliferation is a hallmark of malignancy and may be assessed by a variety of methods, including counting mitotic figures in stained tissue sections, incorporation of labeled nucleotides into DNA, and flow cytometric evaluation of the fraction of the cells in S phase (1–3). The most widely practiced measurement involves the immunohistochemical (IHC) assessment of Ki67 antigen (also known as antigen identified by monoclonal antibody Ki-67 [MKI67]), a nuclear marker expressed in all phases of the cell cycle other than the G<sub>0</sub> phase (4). In spite of consistent data on Ki67 as a prognostic marker in early breast cancer, its role in breast cancer management remains uncertain (5). As shown by Urruticoechea et al. (6), 17 of the 18 studies that included more than 200 patients showed statistically significant association between Ki67 and prognosis providing compelling evidence for a biological relationship, but the cutoffs to distinguish “Ki67 high” from “Ki67 low” varied from 1% to 28.6%, thereby severely limiting its clinical utility.

On March 12, 2010, investigators representing translational research efforts from many of the cooperative breast cancer groups in both North America and Europe were convened by Dr Dowsett and Dr Hayes, respective cochair of the Breast International Group and North American Breast Cancer Group Biomarker Working Party, at the Breakthrough Breast Cancer Research Centre (London) to review the present state of the art of Ki67 evaluation and its potential utility. These investigators, designated

the “International Ki67 in Breast Cancer Working Group,” agreed that Ki67 measurement by IHC was the current assay of choice for measuring and monitoring tumor proliferation in standard pathology specimens. However, they recognized the poor agreement on the precise clinical uses of Ki67 and the substantial heterogeneity and variable levels of validity in methods of assessment.

In this study, the International Ki67 in Breast Cancer Working Group proposed guidelines for the analysis, reporting, and use of Ki67 that should reduce interlaboratory variability and improve interstudy comparability of Ki67 results. Some issues cannot be fully resolved at this stage because of limited evidence to make a firm recommendation. Nonetheless, following this guidance should enable improved comparison and pooling of data and more rapid establishment or rejection of the utility of Ki67 in breast cancer management.

The goals of this study were 1) to provide an account of the substantive data that have identified a potentially valuable clinical role for Ki67 measurement; this is reported in a concise manner because of the availability of a recent detailed review (5); 2) to consider the methodological variables that influence the measurement of Ki67 and often result in lack of analytical validity; and 3) to offer guidelines, based on current evidence, that should allow harmonization of methodology and, we hope, lead to the definition of the clinical utility of this potentially important marker.

## Roles of Ki67 in Clinical Management and Research: Clinical Utility

### Adjuvant Therapy

**Prognostic Role of Ki67.** Many studies have demonstrated the prognostic value of Ki67 (5); however, almost all studies are retrospective, and many include heterogeneous groups of patients who were treated and followed in various ways that are often incompletely documented. Furthermore, the assays for Ki67 were performed with different methods, and cutoffs to designate “positive” and “negative” or “high” and “low” Ki67 populations differ widely. As a result, the American Society of Clinical Oncology (ASCO) Tumor Marker Guidelines Committee determined that the evidence supporting the clinical utility of Ki67 was insufficient to recommend routine use of this marker for prognosis in patients with newly diagnosed breast cancer (7).

The clinical utility of Ki67 as a prognostic marker might be more apparent if it were considered within more narrowly defined tumor subgroups and/or as part of a multiparameter panel of biomarkers. For example, investigators have generated an IHC-based assay of four markers, designated IHC4, which consists of estrogen receptor (ER), progesterone receptor (PgR), HER2, and Ki67 (8). Other investigators have reported that Ki67 is an important part of a prognostic algorithm for residual risk in early breast cancer patients treated with letrozole or tamoxifen (9). These results require further analytical and clinical validation before widespread application.

**Predictive Role of Ki67.** Penault-Llorca et al. (10) recently reported that high levels of Ki67 were predictive of benefit from adding docetaxel to fluorouracil and epirubicin chemotherapy as adjuvant treatment for patients with ER-positive tumors in the PACS01 randomized trial. Similar results were seen in the Breast Cancer International Research Group 001 trial (11). The results contrast, however, with those from International Breast Cancer Study Group Trials VIII and IX that found no predictive value of Ki67 levels for the addition of cyclophosphamide, methotrexate, and fluorouracil to endocrine therapy in endocrine-responsive node-negative disease (12). Thus, the data on the identification of patients benefiting from chemotherapy require confirmation before the use of Ki67 reaches clinical utility.

There are fewer data to suggest that Ki67 predicts adjuvant chemotherapy response in ER-negative tumors. Some studies of preoperative chemotherapy, and a few studies of classic adjuvant therapy, strongly suggest that ER-negative tumors as a group are much more responsive to chemotherapy than ER-positive tumors (13,14). Although not confirmed, a straightforward hypothesis is that the higher chemotherapy sensitivity observed in patients with ER-negative tumors is because of the consistently higher rates of proliferation of these tumors. If so, Ki67 levels may be helpful to identify those patients most likely to benefit from chemotherapy (15).

### Neoadjuvant Therapy and Pharmacodynamic Role of Ki67

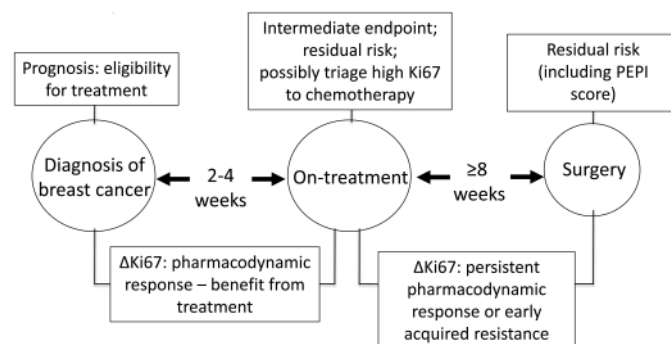
**Background of Neoadjuvant Systemic Therapy.** The administration of systemic therapy before surgery, otherwise designated neoadjuvant or preoperative therapy, offers improvements in surgical outcomes and the opportunity to assess the response of the primary tumor using clinical, biochemical, or molecular markers of

benefit. Because of its well-established role in downstaging disease before surgery, systemic therapy has become a favored clinical trial scenario for the evaluation of novel therapies.

**Neoadjuvant Endocrine Treatment.** In the case of chemotherapy, demonstration of pathological complete response is a validated predictor of disease-free and overall survival. As detailed below, emerging evidence suggests that Ki67 measurement can have several valuable roles (Figure 1).

**Ki67 as an End-of-Neoadjuvant-Treatment Endpoint.** The strongest evidence to support Ki67 as the primary endpoint of neoadjuvant endocrine comparisons is derived from two trials—the Immediate Preoperative Anastrozole, Tamoxifen, or Combined with Tamoxifen (IMPACT) study, comparing neoadjuvant anastrozole vs tamoxifen vs combination of anastrozole and tamoxifen (16), and the P024 study of neoadjuvant letrozole vs tamoxifen (17). In each study, the difference in the degree of Ki67 suppression between the study arms mirrored the difference in recurrence in equivalent large adjuvant trials, Arimidex, Tamoxifen Alone or Combined (ATAC) trial and Breast International Group (BIG) 1-98 trial, respectively (18,19). Similar data have emerged from the neoadjuvant study American College of Surgeons Oncology Group (ACOSOG) Z1031 (20) showing no difference in Ki67 suppression between exemestane and anastrozole, which is in agreement with the results of the MA.27 trial where similar rates of disease-free survival were observed in patients treated with the same agents as adjuvant therapy (21).

Based on these results, and similar observations with Ki67 measured after 2 weeks of endocrine treatment (described below), Ki67 has been used as a primary endpoint in several short-term, “window-of-opportunity” presurgical studies, mainly, but not exclusively, of endocrine treatment (22–24). In addition, in one therapeutic neoadjuvant trial that tested the activity of gefitinib when added to anastrozole (25), Ki67 was chosen to be the primary endpoint, replacing the conventional clinical endpoint of tumor shrinkage. This trial showed no benefit from gefitinib for either Ki67 or clinical response, contributing to the decision not to proceed to test the combination in phase III clinical trials in patients with early breast cancer.



**Figure 1.** Schematic representation of the applications of Ki67 as a pharmacodynamic marker in endocrine neoadjuvant therapy. The circles represent the tumor at different time points during neoadjuvant treatment. The boxes contain the application of Ki67 or of change in Ki67 ( $\Delta$ Ki67) between the respective time points. PEPI = Preoperative Endocrine Prognostic Index.

In the P024 study (17), after 4 months of neoadjuvant endocrine therapy with either letrozole or tamoxifen, the authors observed that Ki67, pathological tumor size, node status, and ER status were independently associated with recurrence-free and overall survival. A Preoperative Endocrine Prognostic Index (PEPI) derived from a combination of these factors was validated as predictive of long-term outcome in an independent dataset from the IMPACT trial (26). As shown by Ellis et al. (26), the PEPI identifies a group of patients at the end of neoadjuvant endocrine therapy with such extremely low risk of recurrence on endocrine therapy alone that they might be spared additional chemotherapy. These authors have suggested that high PEPI scores identify those who most likely should receive chemotherapy, given that their tumors are relatively resistant to endocrine treatment.

**Ki67 as a Pharmacodynamic Intermediate Endpoint.** Absence of a decrease in Ki67 early in treatment might be predictive of therapeutic failure. For example, in the IMPACT trial (16), the value of Ki67 after 2 weeks of endocrine therapy had a stronger association with time to recurrence compared with pretreatment Ki67 level; moreover, association between pretreatment Ki67 level and time to recurrence was not statistically significant in a multivariable model that included both the pretreatment and 2-week Ki67 values (27). Given that the 2-week value results from the pretreatment value, which has prognostic importance, and the change over 2 weeks, which has predictive importance, this observation suggests that the 2-week value integrates both these effects and thereby provides an index of the residual risk after endocrine therapy. The possible advantage of measuring 2-week Ki67 instead of pretreatment Ki67 is under evaluation in the 4000-patient Peri-Operative Endocrine Therapy for Individualizing Care (POETIC) window-of-opportunity study (28).

Using tumor samples accrued from a phase II neoadjuvant trial with letrozole (29), Ellis et al. (26) identified a group of patients in whom the proportion of tumor cells positive for Ki67 was 10% or greater after 4 weeks. As predicted from this relatively high on-treatment value, these patients were very unlikely to be in the PEPI zero category (defined by pathological tumor size  $\leq 5$  cm, node negative, Ki67  $\leq 2.7\%$ , and ER  $>2$  Allred score after endocrine treatment) for which treatment without chemotherapy could be considered. Taken together with the previously published results (26,27), these data suggest that Ki67 evaluation at an early time point can be used to triage ER-positive patients away from neoadjuvant endocrine therapy to neoadjuvant chemotherapy. These investigators are prospectively validating this finding in an extension of the Z1031 trial (cohort b; trial registration number NCT00265759).

**Ki67 as an Eligibility Criterion for Neoadjuvant Trials.** One main objective of many neoadjuvant trials is to provide evidence of activity of a new therapeutic agent. If Ki67 reduction is to be used as a pharmacodynamic or primary endpoint, then patients whose tumors have relatively low Ki67 at diagnosis are unlikely to be informative because they have little potential to be suppressed. It is also unlikely that such patients could benefit from additional therapy, even if it were predicted to work, because of their excellent prognosis. Proposals have therefore emerged that these patients should be excluded from such trials.

**Neoadjuvant Chemotherapy.** Currently, the value of Ki67 during neoadjuvant chemotherapy is less obvious than with neoadjuvant endocrine therapy. Reductions in Ki67 occur in the tumors of most patients receiving neoadjuvant chemotherapy, and there is some evidence that there are greater reductions in patients who respond to treatment (30). A recent study also reported that in patients not having a pathological complete response, Ki67 levels in the residual tumor were strongly associated with outcome (31). This approach is therefore attractive for identifying patients for trials of additional adjuvant therapy after neoadjuvant chemotherapy; such patients stand to benefit most from added therapy, and the high event rate should provide a rapid result.

## Methodological Issues in Ki67 Measurement: Analytical Validity

The above scenarios highlighting areas in which Ki67 measurement may well have clinical utility prompt a need for reproducible methodology and consistent scoring methods; in other words, the analytical validity, as defined by Evaluation of Genomic Applications in Practice and Prevention (EGAPP) (32), needs to be standardized.

Ki67 measurement by IHC has been adopted by many groups because of its particularly favorable biological expression patterns and analytical robustness relative to other biomarkers detected by IHC assays. Nevertheless, there are many steps that introduce variability in the results of these assays. We provide guidance on preferred methodologies to minimize the variability and recommend specific actions to harmonize Ki67 scoring and reporting.

### Preanalytical Validity

Several preanalytical issues might adversely affect Ki67 measurement. These include type of biopsy, time to fixative, type of fixative, time in fixative, and how the specimen is stored long term (Table 1). Data from two recent studies (33,34) suggest that, in general, Ki67 has better tolerance of typical preanalytical variability than most breast cancer IHC assays. For example, in one of these studies (33), Ki67 staining in core-cut biopsies performed on fresh surgical excisions did not vary over 20–80 minutes delay in fixation nor from measurements of whole sections from the same resection specimen. However, differences in the appearance of stained nuclei were frequently apparent in these studies: the more rapidly fixed core-cuts consistently showed well-circumscribed uniformly staining nuclei, whereas nuclei in whole sections often showed areas of highly variable staining (Figure 2). The difference in nuclear integrity between the two staining methods is clear in this figure. This variability did not disrupt the scores derived by visual assessment but can be difficult to deal with in digital image analysis procedures.

Several studies including a systematic interlaboratory and interobserver reproducibility study for IHC assessment of Ki67 found that the following preanalytical factors decrease Ki67 labeling index and should therefore be avoided (35): overnight delay before fixation, freezing the specimen for frozen section analysis before fixation, use of ethanol or Bouin solution rather than neutral buffered formalin fixation, and use of EDTA or acid decalcification protocols (35,36).

**Table 1.** Factors that may affect Ki67 immunohistochemistry\*

Setting	Factor	Variables	Important?	Comments
Preanalytical	Type of biopsy	Core vs whole section	No	Both are suitable. Some data suggest that whole section may give higher scores than core biopsy.
	Type of fixative	Previously frozen, or EtOH or EDTA fixative, or previous acid decalcification vs neutral buffered formalin	Yes	Avoid all but neutral buffered formalin. Others reduce Ki67 staining compared with neutral buffered formalin.
	Time to fixation	Integrity of nuclei	Yes	For visual analysis, has little impact unless extreme. Important for image analysis.
	Means of storage	Tissue in paraffin block vs cut section	Yes	Prolonged storage of formalin-fixed paraffin-embedded tissue block at room temperature has little effect on Ki67. Avoid prolonged exposure to air of cut sections on glass slides.
Analytical	Antigen retrieval	Yes vs no	Yes	Required. Microwave processing recommended.
	Specific antibody	MIB1 vs other antibodies against Ki67 antigen	Yes	MIB1 is the most widely validated antibody. SP6 antibody against Ki67 appears promising but insufficient data to support routine use at this time.
	Colorimetric detection system	Avidin–biotin immunoperoxidase vs polymer detection†	No	Both suitable.
	Counterstain	Completeness and intensity of stain	Yes	Important that all negative nuclei are counterstained.
Interpretation and scoring	Method of reading	Cellular component, staining intensity	Yes	1) Count all positive cells within region in which all nuclei have been stained. 2) Scoring requires determination of percentage cells positive. 3) No interpretation of intensity.
	Area of slide read	Edge vs central; hot spots vs area without hot spots vs all areas	Yes	Controversial: currently recommend average score across the section.
Data analysis	Image	Visual vs automated analysis	Unknown	Unknown whether either method is superior.
	Cut point	Any vs no staining; arbitrary vs data-derived cut point; or continuous variable	Controversial	It is controversial because there is no recommended consensus cut point at this time. Select cut point based on context (prognosis, prediction of specific therapy, selection of patients for trial, use as pharmacodynamic or endpoint biomarker). Endpoint must be validated in separate independent study of similar design with same endpoints.

\* EtOH = ethanol.

† The polymer detection method uses polymeric antibodies and increases the number of available enzymes or ligands binding at the antigenic site, thus increasing their reactivity to chromogen.

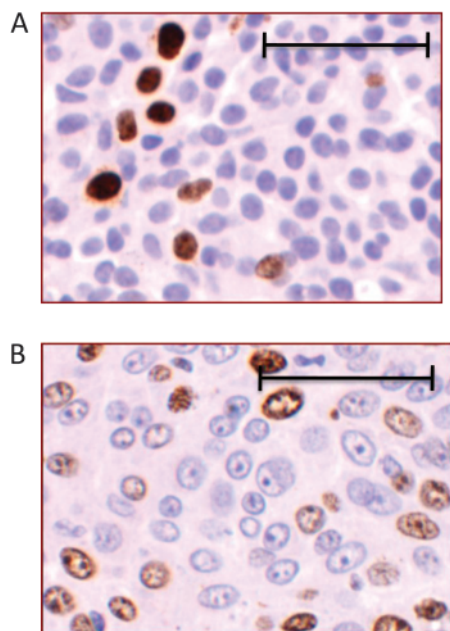
Fixation with neutral buffered formalin for 4–48 hours has been shown to be adequate (37), and fixation even for 154 days was reported to not reduce Ki67 staining substantially (38). Thus, when tissue is fixed in neutral buffered formalin, IHC for Ki67 is robust across a wide range of fixation times. Tissue handling guidelines that are already in place for ER (8–72 hours of neutral buffered formalin fixation) are therefore more than adequate for Ki67 (39,40).

Once tissue is properly fixed and embedded in paraffin, antigenicity is well preserved, potentially for decades (41,42). However, there is a documented loss of Ki67 immunoreactivity if blocks are cut and sections are stored on glass slides exposed to room air for 3 months or longer (43). Paraffin coating of the slide and/or

storage under nitrogen desiccation appears to protect only marginally against loss of antigenicity. Typical room temperature and air storage for up to 2 weeks, however, has no perceptible impact on Ki67 positivity (T. Nielsen, unpublished data).

### Analytical Validity

The detailed characteristics of assays for Ki67 are critical to their results. The original Ki67 antibody was applicable for IHC only in fresh frozen material. Later, with the development of heat-induced epitope retrieval methodologies, mouse monoclonal antibodies were developed with robust and reproducible results in formalin-fixed paraffin-embedded sections. The most commonly used mouse anti-human Ki67 monoclonal antibody, MIB1 clone (42),



**Figure 2.** Examples of Ki67 staining in breast cancer. Tumor biopsies were fixed in neutral buffered formalin and sections stained for Ki67 with the MIB1 antibody (brown stain) and counterstained with Mayer's hematoxylin (blue stain). **A)** Well-fixed specimen. **B)** Poorly fixed specimen. The micrograph was taken using a Leica Microsystems Ariol image analyzer (Leica Microsystems, Gateshead, UK). Scale bar = 50  $\mu$ m.

has the especially favorable property of detecting an epitope motif unique to Ki67 (ensuring specificity) that is repeated 16 times in the protein (enhancing sensitivity) (44). A related advantageous property of MIB1 as a reagent for IHC is its consistent and much better performance across a wide range of antibody dilution and conditions (45) compared with other proliferation markers such as proliferating cell nuclear antigen (PCNA). Although Ki67 IHC is tolerant to a variety of epitope retrieval protocols, protease and low pH methods should be avoided (46).

Given the long and highly validated track record for monoclonal antibody MIB1, we recommend it be considered a “gold standard” against which other antibodies or methods of proliferation analysis should be compared. However, other anti-Ki67 antibodies have been reported which may provide additional incremental advantages. For example, the rabbit anti-human Ki67 monoclonal antibody SP6 (which recognizes the same repeated Ki67 epitope as MIB1) may provide further improvements in sensitivity (47) and in quantitative image analysis (48), and this reagent has been used successfully in several recent studies (49,50).

Chromogen development and counterstaining for Ki67 IHC appear no different than for other antibody–antigen systems. The chromogenic staining is normally very clear, but the degree of counterstaining is important to optimize, given that negative nuclei determine the overall population for calculating the proportion of Ki67-positive cells. Weak counterstaining can result in overestimation of the Ki67 index (51).

### Interpretation of Ki67 Staining and Scoring

Ki67 is a nuclear protein. Cytoplasmic staining and occasionally membrane staining of Ki67 can occur with MIB1 antibody, especially in breast cancer showing squamous metaplastic changes (52).

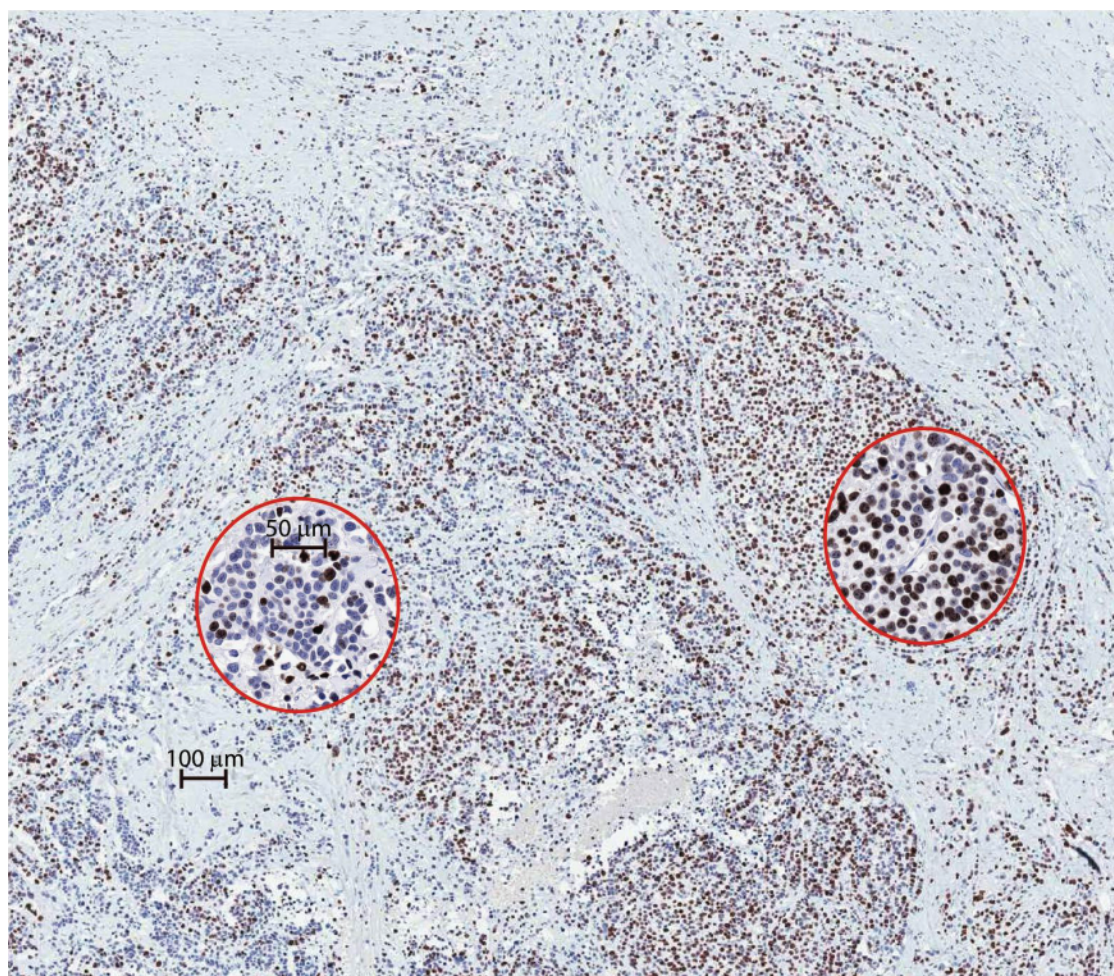
Attention to preanalytical protocols and/or use of SP6 antibody may decrease this extraneous staining to some extent, but when present they should be ignored while creating a Ki67 score. Only nuclear staining (plus mitotic figures which are stained by Ki67) should be incorporated into the Ki67 score that is defined as the percentage of positively stained cells among the total number of malignant cells scored. As with other IHC stains, it is helpful to have internal positive controls: mitotic figures, normal ducts, and lymphocytes as well as, to a lesser extent, endothelial cells and stromal cells serve for this purpose.

If the staining is homogenous, the recommendation is to count at least three randomly selected high-power ( $\times 40$  objective) fields. However, biological heterogeneity of Ki67 staining can occur across specimens, and the location and extent of the area of the cancer that should be scored is controversial. Two types of heterogeneity are prominent: a gradient of increasing staining toward the tumor edge and hot spots. For the former, three fields should be scored at the periphery of the tumor because the invasive edge is widely considered to be the most biologically active part of the tumor and is most likely to drive outcome of the disease. An exception to this recommendation is, if comparisons are to be made between Ki67 staining on whole sections with those from core-cuts, for example, core cuts taken in presurgical studies. Preferably core cuts taken at surgery would be used for such comparisons, but if this is not possible, then scoring of the excision should involve fields from across the whole tumor and not just the periphery.

Hot spots, defined as areas in which Ki67 staining is particularly prevalent, may occur in an otherwise homogeneously stained sample (Figure 3). The Ki67 score would be approximately 30% for the circled area on the left and approximately 90% for the circled area on the right in this figure. The approach to scoring hot spots varies across studies; some investigators have focused in particular on the analysis of hot spots, others have included hot spots in a general assessment of Ki67 across the section, and yet others have recommended avoiding them altogether. This issue needs clarification, and a working party of the International Ki67 in Breast Cancer Working Group has been established to assess which method is more robust. In the meantime, for the purposes of consistency, when hot spots are present, an approach that assesses the whole section and records the overall average score is recommended.

Mostly, between 500 and 2000 tumor cells have been scored in published studies. Core-cut biopsies are most frequently used for diagnostics these days (as recommended by ASCO/College of American Pathologists [CAP] for ER and PgR) (39,40) and for research studies in which Ki67 acts as a dynamic marker; all the invasive tumor cells can be scored in such samples. However, where scoring all cells is impractical, to achieve adequate precision, we recommend the interpreting pathologist scores at least 1000 cells and that 500 cells be accepted as the absolute minimum. These cell numbers should be scored in fields that are seen to be representative on an initial overview of the whole section.

Tissue microarrays (TMAs) are an increasingly popular and influential resource for assessing the relationship of biomarkers, including Ki67, with outcome in large phase III clinical trials or epidemiological studies. There are no published systematic comparisons of the assessment of Ki67 on TMAs vs whole sections in breast cancer, but there is anecdotal evidence that scores are



**Figure 3.** Variable levels of Ki67 staining in breast cancer. Tumor biopsies were fixed in neutral buffered formalin and sections stained for Ki67 with MIB1 antibody (**brown stain**) and counterstained with Mayer's hematoxylin (**blue stain**). The two areas **circled in red** are shown at higher magni-

fication to illustrate the differences in scores that can occur in different high-power fields. The average score across the whole section should be taken. The micrograph was taken using Aperio ScanScope image analyzer (Aperio, Vista, CA). Scale bar = 100  $\mu$ m and scale bar of inset = 50  $\mu$ m.

generally lower on TMAs. Until data assessing the relationship between TMA scores and clinical samples are published, Ki67 studies in TMAs should not be used for setting quantitative relationships or establishing cutoffs for clinical application on other types of samples.

Most data in the literature are derived from visual scoring, which may be aided by the use of a grid. Digital imaging may be helpful, but because all stained malignant cells are regarded as positive, irrespective of the intensity of stain, the contribution of imaging to removal of subjective bias is less important for Ki67 than with some markers (eg, ER, HER2). As noted above, the loss of integrity of the interior of nuclear material may make the selection of positive nuclei more difficult for some image analysis systems.

### Data Handling

Ki67 measurements generally follow a log-normal distribution [eg, see Jones et al. (15)]. Summary statistics and comparative analytical methods should be based on log-transformed Ki67 data, or alternatively on nonparametric methods.

Methods to develop cut points to distinguish positive from negative or high from low tumor marker results have been widely

discussed in the literature (53). For IHC of Ki67, many cutoffs have been used, although staining levels of 10%–20% have been the most common to dichotomize populations (54). However, without standardization of methodology, these cutoffs have limited value outside of the studies from which they were derived and the centers that performed them. This issue is also context related: A threshold that is appropriate for determination of prognosis may not pertain to one that is used for eligibility for a neoadjuvant trial or for use of Ki67 as a pharmacodynamic marker. Currently, in the absence of harmonized methodology, the International Ki67 in Breast Cancer Working Group was unable, therefore, to come to consensus regarding the ideal cut point(s) that might be used in clinical practice.

Changes in levels of Ki67 when used as a pharmacodynamic marker in window-of-opportunity or neoadjuvant trials have been most frequently expressed as a percentage of the baseline value, but there are few, if any, validated data to demonstrate precisely what percentage change is clinically important. Changes can also be problematic to determine if baseline values are very low. The International Ki67 in Breast Cancer Working Group identified better definition of a meaningful change in Ki67 as an important research question.

## Conclusion and Recommendations

Overall, the International Ki67 in Breast Cancer Working Group concluded that measures of proliferation could be important both in standard clinical practice and, particularly, within clinical trials. Of these, Ki67, as assessed by IHC with monoclonal antibody MIB1, has the largest body of literature support. Although preanalytical and analytical issues affect its measurement Ki67 is one of the most robust biomarkers measured by IHC, showing relatively consistent measurements in specimens across a range of conditions used in routine fixation, tissue processing, and IHC analysis. Scoring procedures however vary at present, and their lack of standardization for different types of specimens (eg, core-cuts vs whole-tumor sections vs TMAs) is problematic. Perhaps, equally importantly no established quality assurance schemes are in place to ensure that the procedures for Ki67 analysis in one laboratory lead to scores comparable to those in others. Thus, the direct

application of specific cutoffs for decision making must be considered unreliable unless analyses are conducted in a highly experienced laboratory with its own reference data. The same issues prohibit comparisons of Ki67 data between clinical trials.

To drive forward harmonization, we have initiated a pilot between-laboratory quality assessment schemes. We aim to extend these to all interested researchers and also to create TMAs with consensus scores that can be used for standardization by those new to the field to standardize their procedures. We also propose that access to large tissue collections from adjuvant trials should be welcomed for Ki67 analysis when such analysis applies these standardization and quality assurance (ie, QA) materials and adheres to the recommendations in this report. Further studies of scoring methodology are also underway, and data from these will be published. The results of these initiatives may lead to some future clarifications in our recommendations, which are presented below (Box 1).

### Box 1. Recommendations for Ki67 assessment in breast cancer

#### Preanalytical

- Core-cut biopsies and whole sections from excision biopsies are acceptable specimens; when comparative scores are to be made, it is preferable to use the same type for both samples (eg, in presurgical studies).
- TMAs are acceptable for clinical trial evaluation or epidemiological studies of Ki67.
- Fixation in neutral buffered formalin should follow the same guidelines as published for steroid receptors (39,40).
- Once prepared, tissue sections should not be stored at room temperature for longer than 14 days. Results after longer storage must be viewed with caution.

#### Analytical

- Known positive and negative controls should be included in all batches; positive nuclei of nonmalignant cells and with mitotic figures provide evidence of the quality of an individual section.
- Antigen retrieval procedures are required. The best evidence supports the use of heat-induced retrieval most frequently by microwave processing.
- The MIB1 antibody is currently endorsed for Ki67.

#### Interpretation and scoring

- In full sections, at least three high-power (×40 objective) fields should be selected to represent the spectrum of staining seen on initial overview of the whole section.
- For the purpose of prognostic evaluation, the invasive edge of the tumor should be scored.
- If pharmacodynamic comparisons must be between core cuts and sections from the excision, assessment of the latter should be across the whole tumor.
- If there are clear hot spots, data from these should be included in the overall score.
- Only nuclear staining is considered positive. Staining intensity is not relevant.
- Scoring should involve the counting of at least 500 malignant invasive cells (and preferably at least 1000 cells) unless a protocol clearly states reasons for fewer being acceptable.
- Image analysis methods for Ki67 remain to be proven for use in clinical practice.

#### Data handling

- The Ki67 score or index should be expressed as the percentage of positively staining cells among the total number of invasive cells in the area scored.
- Statistical analysis should take account of the log-normal distribution generally followed by Ki67 measurement.
- The most appropriate endpoint in comparative studies of treatment efficacy or response is the percentage suppression of Ki67-positive cells.
- The most appropriate endpoint for assessing residual risk of recurrence is the on-treatment proportion of Ki67-positive cells.
- Cut points for prognosis, prediction, and monitoring should only be applied if the results from local practice have been validated against those in studies that have defined the cutoff for the intended use of the Ki67 result.

## References

1. Tubiana M, Pejovic MH, Chavaudra N, Contesso G, Malaise EP. The long-term prognostic significance of the thymidine labelling index in breast cancer. *Int J Cancer*. 1984;33(4):441–445.
2. Dressler LG, Seamer L, Owens MA, Clark GM, McGuire WL. Evaluation of a modeling system for S-phase estimation in breast cancer by flow cytometry. *Cancer Res*. 1987;47(20):5294–5302.
3. Tovey SM, Witton CJ, Bartlett JM, Stanton PD, Reeves JR, Cooke TG. Outcome and human epidermal growth factor receptor (HER) 1–4 status in invasive breast carcinomas with proliferation indices evaluated by bromodeoxyuridine labelling. *Breast Cancer Res*. 2004;6(3):R246–R251.
4. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J Immunol*. 1984;133(4):1710–1715.
5. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol*. 2010;11(2):174–183.
6. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol*. 2005;23(28):7212–7220.
7. 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(33):5287–5312.
8. Cuzick J, Dowsett M, Wale C, et al. Prognostic value of a combined ER, PgR, Ki67, HER2 immunohistochemical (IHC4) score and the comparison with the GHI recurrence score—results from TransATAC. *Cancer Res*. 2009;69(suppl 24):503s.
9. Viale G, Regan MM, Dell'Orto P, et al. Which patients benefit most from adjuvant aromatase inhibitors? Results using a composite measure of prognostic risk in the BIG 1–98 randomised trial. *Ann Oncol*. 2011. doi:10.1093.annonc/mdq738.
10. Penault-Llorca F, Andre F, Sagan C, et al. Ki67 expression and docetaxel efficacy in patients with estrogen receptor-positive breast cancer. *J Clin Oncol*. 2009;27(17):2809–2815.
11. Hugh J, Hanson J, Cheang MC, et al. Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *J Clin Oncol*. 2009;27(8):1168–1176.
12. Viale G, Regan MM, Mastropasqua MG, et al. Predictive value of tumor Ki-67 expression in two randomized trials of adjuvant chemoendocrine therapy for node-negative breast cancer. *J Natl Cancer Inst*. 2008;100(3):207–212.
13. Guarneri V, Broglio K, Kau SW, et al. Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors. *J Clin Oncol*. 2006;24(7):1037–1044.
14. Ring AE, Smith IE, Ashley S, Fulford LG, Lakhani SR. Oestrogen receptor status, pathological complete response and prognosis in patients receiving neoadjuvant chemotherapy for early breast cancer. *Br J Cancer*. 2004;91(12):2012–2017.
15. Jones RL, Salter J, A'Hern R, et al. Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer. *Breast Cancer Res Treat*. 2010;119(2):315–323.
16. Dowsett M, Smith IE, Ebbs SR, et al. Short-term changes in Ki-67 during neoadjuvant treatment of primary breast cancer with anastrozole or tamoxifen alone or combined correlate with recurrence-free survival. *Clin Cancer Res*. 2005;11(2, pt 2):951s–958s.
17. Ellis MJ, Coop A, Singh B, et al. Letrozole inhibits tumor proliferation more effectively than tamoxifen independent of HER1/2 expression status. *Cancer Res*. 2003;63(19):6523–6531.
18. Baum M, Budzar AU, Cuzick J, et al. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. *Lancet*. 2002;359(9324):2131–2139.
19. Thurlimann B, Keshaviah A, Coates AS, et al. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N Engl J Med*. 2005;353(26):2747–2757.
20. Ellis MJ, Suman VJ, Hoog J, et al. ACOSOG Z1031, a randomized phase 2 neoadjuvant comparison between letrozole, anastrozole and exemestane for postmenopausal women with ER rich stage 2/3 breast cancer: clinical and biomarker outcomes. *J Clin Oncol*. 2011;29(17):2342–2349.
21. Goss PE, Ingle JN, Chapman J-AW, et al. Final analysis of NCIC CTG MA.27: a randomized phase III trial of exemestane versus anastrozole in postmenopausal women with hormone receptor positive primary breast cancer. *Cancer Res*. 2010;70(24)(suppl 2):75s.
22. Dowsett M, Bundred NJ, Decensi A, et al. Effect of raloxifene on breast cancer cell Ki67 and apoptosis: a double-blind, placebo-controlled, randomized clinical trial in postmenopausal patients. *Cancer Epidemiol Biomarkers Prev*. 2001;10(9):961–966.
23. Guix M, Granja Nde M, Meszoely I, et al. Short preoperative treatment with erlotinib inhibits tumor cell proliferation in hormone receptor-positive breast cancers. *J Clin Oncol*. 2008;26(6):897–906.
24. Robertson JF, Nicholson RI, Bundred NJ, et al. Comparison of the short-term biological effects of 7alpha-[9-(4,4,5,5,5-pentafluoropentylsulfinyl)-nonyl]estra-1,3,5, (10)-triene-3,17beta-diol (Faslodex) versus tamoxifen in postmenopausal women with primary breast cancer. *Cancer Res*. 2001;61(18):6739–6746.
25. Smith IE, Walsh G, Skene A, et al. A phase II placebo-controlled trial of neoadjuvant anastrozole alone or with gefitinib in early breast cancer. *J Clin Oncol*. 2007;25(25):3816–3822.
26. Ellis MJ, Tao Y, Luo J, et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst*. 2008;100(19):1380–1388.
27. Dowsett M, Smith IE, Ebbs SR, et al. Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J Natl Cancer Inst*. 2007;99(2):167–170.
28. Dowsett M, Smith I, Robertson J, et al. Endocrine therapy, new biologicals and new study designs for pre-surgical studies in breast cancer. *J Natl Cancer Inst*. 2011. Submitted.
29. Olson JA Jr., Budd GT, Carey LA, et al. Improved surgical outcomes for breast cancer patients receiving neoadjuvant aromatase inhibitor therapy: results from a multicenter phase II trial. *J Am Coll Surg*. 2009;208(5):906–914; discussion 915–916.
30. Assersohn L, Salter J, Powles TJ, et al. Studies of the potential utility of Ki67 as a predictive molecular marker of clinical response in primary breast cancer. *Breast Cancer Res Treat*. 2003;82(2):113–123.
31. Jones RL, Salter J, A'Hern R, et al. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat*. 2009;116(1):53–68.
32. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med*. 2009;11(1):3–14.
33. Pinhel IF, Macneill FA, Hills MJ, et al. Extreme loss of immunoreactive p-Akt and p-Erk1/2 during routine fixation of primary breast cancer. *Breast Cancer Res*. 2010;12(5):R76.
34. Bai Y, Tolles J, Cheng H, et al. Quantitative assessment shows loss of antigenic epitopes as a function of time to formalin fixation. *Modern Pathol*. 2011;91(8):1253–1261.
35. Mengel M, von Wasielewski R, Wiese B, Rudiger T, Muller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J Pathol*. 2002;198(3):292–299.
36. Benini E, Rao S, Daidone MG, Pilotti S, Silvestrini R. Immunoreactivity to MIB-1 in breast cancer: methodological assessment and comparison with other proliferation indices. *Cell Prolif*. 1997;30(3–4):107–115.
37. Munakata S, Hendricks JB. Effect of fixation time and microwave oven heating time on retrieval of the Ki-67 antigen from paraffin-embedded tissue. *J Histochem Cytochem*. 1993;41(8):1241–1246.
38. Arber DA. Effect of prolonged formalin fixation on the immunohistochemical reactivity of breast markers. *Appl Immunohistochem Mol Morphol*. 2002;10(2):183–186.
39. 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. *Arch Pathol Lab Med*. 2010;134(6):907–922.
40. 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 (unabridged version). *Arch Pathol Lab Med.* 2010; 134(7):e48–e72.

41. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest.* 2000;80(12):1943–1949.
42. Cattoretti G, Becker MH, Key G, et al. Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave-processed formalin-fixed paraffin sections. *J Pathol.* 1992;168(4):357–363.
43. DiVito KA, Charette LA, Rimm DL, Camp RL. Long-term preservation of antigenicity on tissue microarrays. *Lab Invest.* 2004;84(8):1071–1078.
44. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. *J Cell Physiol.* 2000;182(3):311–322.
45. McCormick D, Yu C, Hobbs C, Hall PA. The relevance of antibody concentration to the immunohistological quantification of cell proliferation-associated antigens. *Histopathology.* 1993;22(6):543–547.
46. Boon ME. Microwave-antigen retrieval: the importance of pH of the retrieval solution for MIB-1 staining. *Eur J Morphol.* 1996;34(5):375–379.
47. Wong SC, Chan JK, Lo ES, et al. The contribution of bifunctional SkipDewax pretreatment solution, rabbit monoclonal antibodies, and polymer detection systems in immunohistochemistry. *Arch Pathol Lab Med.* 2007;131(7):1047–1055.
48. Zabaglo L, Salter J, Anderson H, et al. Comparative validation of the SP6 antibody to Ki67 in breast cancer. *J Clin Pathol.* 2010;63(9):800–804.
49. 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(10):736–750.
50. Niu G, Sun X, Cao Q, et al. Cetuximab-based immunotherapy and radio-immunotherapy of head and neck squamous cell carcinoma. *Clin Cancer Res.* 2010;16(7):2095–2105.
51. Turbin DA, Gao D, Garratt J, et al. Effect of nuclear counterstain on Ki67 proliferative index. Abstract ID 23 / #60777–ASCO 2010 Breast Cancer Symposium (Washington DC, October 1–3, 2010).
52. Faratian D, Munro A, Twelves C, Bartlett JM. Membranous and cytoplasmic staining of Ki67 is associated with HER2 and ER status in invasive breast carcinoma. *Histopathology.* 2009;54(2):254–257.
53. Hollander N, Sauerbrei W, Schumacher M. Confidence intervals for the effect of a prognostic factor after selection of an ‘optimal’ cutpoint. *Stat Med.* 2004;23(11):1701–1713.
54. Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast.* 2008;17(4):323–334.

## Funding

Breast Cancer Research Foundation provided funding for the meeting (through a grant to DFH); Royal Marsden National Institute of Health Research

Biomedical Research Centre, Breakthrough Breast Cancer, and Cancer Research UK (to MD); Fashion Footwear Charitable Foundation of New York and QVC Presents Shoes on Sale (grant to DFH).

## Notes

T. Nielsen holds stock in Bioclassifier LLC and C. Sotiriou is co-inventor of the Genomic Grade Index. All other authors declare no conflict of interest.

We are grateful to Allen M. Gown, David L. Rimm, Dmitry Turbin, Doris Gao, Blake Gilks, and Robert Wolber for their sharing of unpublished data, to Samuel Leung for providing the micrographs for Figure 3, and to Leah Kamin for organizational assistance.

The authors are solely responsible for the study design, data collection, analysis and interpretation of the data, writing the article, and decision to submit the article for publication.

**Affiliations of authors:** Academic Department of Biochemistry, Royal Marsden Hospital and Breakthrough Breast Cancer Centre, London, UK (MD, JS); University of British Columbia, Anatomical Pathology JP1401, Vancouver Hospital, Vancouver, BC, Canada (TON); Cancer Research UK Clinical Trials & Statistics Unit, The Institute of Cancer Research, Sutton, Surrey, UK (RAH); Edinburgh Cancer Research Centre, University of Edinburgh, Western General Hospital, Edinburgh, UK (JB); Division of Cancer, Imperial College, Hammersmith Hospital, London, UK (RCC); Wolfson Institute of Preventive Medicine, Bart’s & The London School of Medicine and Dentistry, Queen Mary University of London, London, UK (JC); Breast Cancer Program, Siteman Cancer Center, Washington University School of Medicine, St Louis, MO (ME); Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI (NLH, DFH); Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada (JCH); Diagnostics Evaluation Branch, Cancer Diagnosis Program, National Cancer Institute, Rockville, MD (TL); National Cancer Institute, Biometric Research Branch, DCTD, Bethesda, MD (LMcS); Division of Pathology, NSABP Foundation, Pittsburgh, PA (SP); Department of Biopathology, University of Auvergne, Clermont-Ferrand cedex, France (FP-L); Laboratorio de Patología Molecular, Hospital Universitari Vall d’Hebron, Edificio Anatomía Patológica, Barcelona, Spain (LP); IBCSG Statistical Center, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA (MR); Department of Medical Oncology, Institut Jules Bordet, and Breast Cancer Translational Research Laboratory (BCTL), Brussels, Belgium (CS); Breast Unit, The Royal Marsden Hospital, London, UK (IES); Department of Pathology, European Institute of Oncology, University of Milan, Milan, Italy (GV); Cancer Therapy Evaluation Program, National Cancer Institute, Rockville, MD (JAZ).