Pathology (June 2017) **49(4)**, pp. 371–378

ANATOMICAL PATHOLOGY

Validity and reliability of Ki-67 assessment in oestrogen receptor positive breast cancer

NICOLA JING[1,2], CATHERINE FANG[3] AND DAVID S. WILLIAMS[1,3,4]

[1]*Department of Pathology, Austin Health, Heidelberg,* [2]*Melbourne Medical School, University of Melbourne, Parkville,* [3]*School of Cancer Medicine, Olivia Newton-John Cancer Research Institute, Latrobe University, Heidelberg, and* [4]*Department of Pathology, University of Melbourne, Parkville, Vic, Australia*

Summary

Ki-67 is a prognostic and predictive biomarker in oestrogen receptor positive breast cancer. However, its measurement is not well standardised. This study compared the validity, intra- and inter-observer reproducibility and reporting time of five methods of Ki-67 assessment on tissue microarrays (TMA) and whole slides. Ki-67 labelling index (LI) was assessed on 71 breast carcinomas of no special type (NST), using five methods: manual counting (gold standard), unaided visual estimation, visual estimation aided by reference photographs, semi-manual digital image analysis (DIA) and fully automated DIA (Aperio platform). On TMA, semi-manual DIA demonstrated the closest agreement with the gold standard [intra-class correlation coefficient (ICC)=0.99 (95% confidence interval 0.98–0.99)]. All other methods also demonstrated close agreement [unaided estimation ICC=0.92 (0.90–0.93), aided estimation ICC=0.93 (0.92–0.95), fully automated DIA ICC=0.97 (0.96–0.97)]. On whole slides, both aided estimation and semi-manual DIA demonstrated excellent agreement with the gold standard [aided visual estimation ICC=0.91 (0.85–0.94), semi-manual DIA ICC=0.94 (0.89–0.96)]. Aided visual estimation significantly improved inter-observer reproducibility compared to unaided estimation [unaided ICC=0.87 (0.80–0.92); aided ICC=0.96 (0.93–0.97)] and corrected the underestimation bias seen in unaided estimation. Importantly, validity and reproducibility on whole slides were lower than on TMA for all methods of assessment, suggesting that field selection is an important source of variability in Ki-67 assessment. Values close to clinically used cut-off values therefore should be interpreted with caution.

Key words: Ki-67; proliferation marker; breast cancer; tissue microarrays; digital image analysis; analytical validity; reproducibility of results.

Received 4 October 2016, revised 22 January, accepted 5 February 2017
Available online 24 April 2017

INTRODUCTION

Kiel-67 antigen (Ki-67) is a well-studied prognostic and predictive biomarker in breast cancer. It is a nuclear protein expressed in all phases of the cell cycle except G0, and therefore indicates which cells are proliferating.[1,2] Ki-67 expression is assessed by immunohistochemistry as the Ki-67 labelling index (LI), which refers to the percentage of positively stained tumour cell nuclei.[3] The test is inexpensive and readily available in diagnostic pathology laboratories, allowing for rapid turnaround times to facilitate clinical decision-making.

Numerous clinical applications for Ki-67 have been proposed. The prognostic value of Ki-67 for disease-free survival and overall survival in early oestrogen receptor (ER) positive breast cancer has been consistently demonstrated.[4–6] A role for the prediction and monitoring of neoadjuvant therapy is also emerging, particularly for endocrine therapy.[7–9] The importance of Ki-67 expression has been supported by gene expression profiling studies. For example, expression of *MKI67* (the gene encoding the Ki-67 protein) and other proliferation-associated genes helps to distinguish between the luminal A and luminal B intrinsic subtypes of breast cancer.[10,11] Moreover, *MKI67* expression is included in gene expression profiling-based tests which predict benefit from chemotherapy, such as Oncotype DX, Genomic Grade Index and PAM50.[12–14] However, because of the high cost of these tests (e.g., Oncotype DX costs $4000 in Australia[15]), there is interest in the use of immunohistochemical profiling, including measurement of Ki-67, ER, progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) as cost-effective surrogate markers for identifying intrinsic subtypes and calculating recurrence risk.[16,17]

However, there is concern about the analytical validity of Ki-67,[18] and there is a lack of consensus regarding its measurement. Variability in pre-analytic, analytic and scoring protocols makes it difficult to implement the cut-off values for clinical decision-making proposed in the literature.[1–3,19] Furthermore, the relationship between Ki-67 LI assessed on whole slides and tissue microarrays (TMA) is not well studied, despite the widespread use of cut-off values established by studies which assessed Ki-67 LI on TMA.[16] Although the International Ki-67 in Breast Cancer Working Group has published consensus guidelines for the assessment of Ki-67,[3] these recommendations have not been widely implemented.[20] Of note, the proposed gold standard is manual counting of at least 1000 cells at high power,[3] which is labour intensive and may imply a false sense of precision. Visual estimation has been proposed as a rapid alternative, but its validity and reliability are disputed.[21–23] Digital image analysis (DIA) is emerging as a highly reproducible technique, but is not yet widely adopted.[21–25]

We performed a concordance study to compare five different methods of Ki-67 assessment, including the gold

Print ISSN 0031-3025/Online ISSN 1465-3931
DOI: http://dx.doi.org/10.1016/j.pathol.2017.02.001

standard and different methods of visual estimation and DIA. We assessed validity, intra- and inter-observer reproducibility and reporting time to determine which is most appropriate for use at our institution. We also compared the use of TMA with whole slides.

MATERIALS AND METHODS

Ethics and patients

The study was performed on a random series of 71 invasive breast cancers diagnosed at the Austin Hospital, Melbourne, Australia. Data were extracted from the Kestral pathology database to identify cases of breast cancer diagnosed between 1 January 2005 and 12 December 2010. Eligibility criteria were: invasive carcinoma of no special type, ER positive. Tumour grade (Elston-Ellis modification of Scarff-Bloom-Richardson grade; BRE), ER, PR and HER2 characteristics were extracted from routine reports. The study was approved by the Austin Hospital Human Research Ethics Committee.

Tissue microarray construction

TMAs were constructed using a Mark II TMA arrayer (Beecher Instruments, USA). Six 1.0 mm cores were selected from each tumour (three from the centre and three from the periphery). Sections (4 μm thickness) were then cut from array blocks and transferred to glass slides for immunohistochemical analysis.

Immunohistochemistry

Immunohistochemistry for Ki-67 was performed by Benchmark Ultra (Ventana, USA). Antigen retrieval was done by heat retrieval in CC1 buffer (Ventana) at 95°C for 52 min. The primary antibody SP6 (Cell Marque, USA) was diluted 1:100 and applied for 32 min at 36°C. Bound primary antibody was detected using UltraView Universal DAB detection kit (Ventana). Slides were counter-stained using haemotoxylin.

Digital image analysis

Slides were scanned using Aperio ScanScope XT (Leica Biosystems, Germany). Parameters for fully automated use of Aperio ImageScope's Nuclear Algorithm were visually tuned on a small number of randomly selected cases. As the fully automated algorithm had only moderate accuracy in identifying tumour cells and distinguishing them from stromal cells, a second semi-manual algorithm was developed using positive selection, which required user input to manually draw around groups of tumour cells (Fig. 1).

Assessment of Ki-67

Ki-67 LI was assessed on TMA and whole slides by a trained medical student and a breast pathologist. Quality control was performed with review of cases by a breast pathologist to ensure that only invasive tumour cells were counted. The following methods were used:

1. Gold standard: manual counting of at least 500 cells on TMA or at least 1000 cells on whole slides, using the Cell Counter plugin for ImageJ. At least three high power fields were selected. Invasive tumour cells were marked blue (any degree of nuclear staining was considered positive) or red (unstained).
2. Unaided visual estimation at high power. Ki-67 was estimated as a range and the midpoint taken for analysis.
3. Aided visual estimation at high power using printed reference images demonstrating Ki-67 LI=5%, 10%, 15%, 20%, 30%, 50%. Reference images were chosen from micrographs counted using the gold standard and are presented in Supplementary Fig. 1–12 (Appendix A).
4. Semi-manual DIA (Aperio), requiring user input to draw around tumour cells, specifically excluding stroma.
5. Fully automated DIA (Aperio), requiring user input to select a field of interest.

Each observation was made blinded to previous observations and clinicopathological data. Measurements were repeated by the same observer using unaided estimation, aided estimation and semi-manual DIA. Measurements on whole slides were also repeated by a second observer (breast pathologist) using manual counting (17 cases) where the same fields were assessed, semi-manual and fully automated DIA (17 cases, with independent field selection and parameter selection), and unaided estimation (all cases, independent field selection) and aided estimation (all cases, independent field selection).

Following the International Ki-67 in Breast Cancer Working Group guidelines, Ki-67 was assessed on whole slides at a minimum of three high power fields at the invasive front.[3] Where hotspots (i.e., areas of high Ki-67 expression) were present, the whole slide average was assessed.[3] For manual counting, at least 1000 cells were counted. The mean time to report the Ki-67 LI for one slide was recorded.

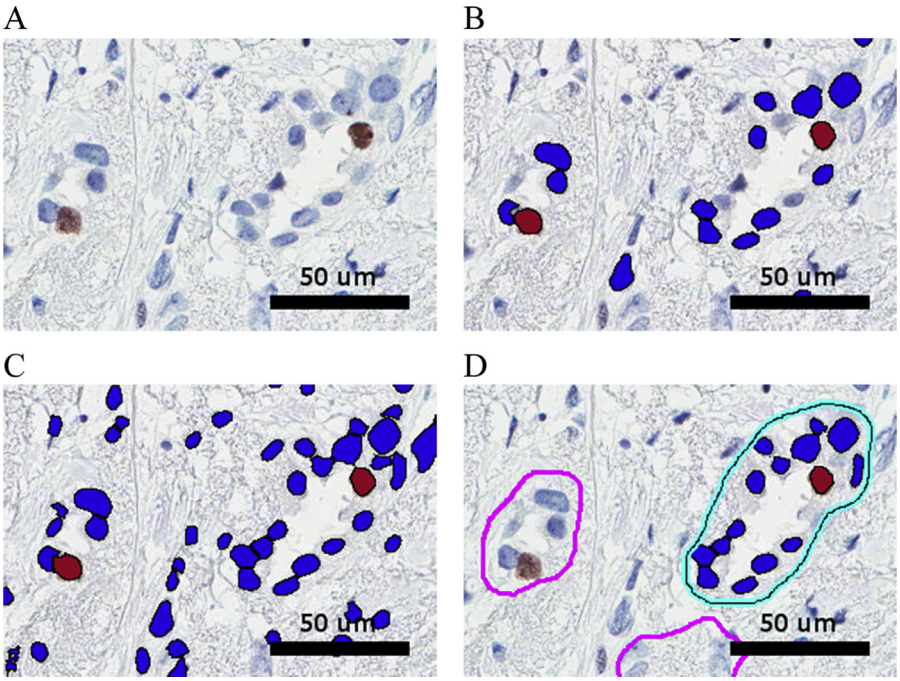

Fig. 1 Examples of the use of digital image analysis (Aperio) for determination of Ki-67 labelling index. (A) Invasive breast carcinoma of no special type, immunohistochemistry for Ki-67. (B) Markup image using fully automated digital image analysis (DIA) on Aperio platform. Blue, negatively stained nuclei; brown, positively stained nuclei. Note that some tumour cells have failed to be detected, and some stromal cells have been mistakenly included. (C) DIA performed using a second algorithm on the Aperio platform. Whilst all tumour cells are detected using this algorithm, more stromal cells have also been mistakenly included. (D) Semi-manual DIA performed using the same algorithm as in C, but clusters of tumour cells have been circled manually to exclude stromal cells from the analysis.