

# Prognostic value of automated KI67 scoring in breast cancer: a centralised evaluation of 8,088 patients from 10 study groups.

## Authors

Mustapha Abubakar<sup>1</sup> ([Mustapha.Abubakar@icr.ac.uk](mailto:Mustapha.Abubakar@icr.ac.uk)); Nick Orr<sup>2</sup> ([Nicholas.Orr@icr.ac.uk](mailto:Nicholas.Orr@icr.ac.uk)); Frances Daley<sup>2</sup> ([Frances.Daley@icr.ac.uk](mailto:Frances.Daley@icr.ac.uk)); Penny Coulson<sup>1</sup> ([Penny.Coulson@icr.ac.uk](mailto:Penny.Coulson@icr.ac.uk)); H Raza Ali<sup>3</sup> ([Raza.Ali@cruk.cam.ac.uk](mailto:Raza.Ali@cruk.cam.ac.uk)); Fiona Blows<sup>4</sup> ([fmb28@medschl.cam.ac.uk](mailto:fmb28@medschl.cam.ac.uk)); Javier Benitez<sup>5,6</sup> ([jbenitez@cnio.es](mailto:jbenitez@cnio.es)); Roger Milne<sup>7,8</sup> ([Roger.Milne@cancervic.org.au](mailto:Roger.Milne@cancervic.org.au)); Herman Brenner<sup>9,10,11</sup> ([h.brenner@dkfz.de](mailto:h.brenner@dkfz.de)); Christa Stegmaier<sup>12</sup> ([c.stegmaier@soziales.saarland.de](mailto:c.stegmaier@soziales.saarland.de)); Arto Mannermaa<sup>13,14</sup> ([arto.mannermaa@uef.fi](mailto:arto.mannermaa@uef.fi)); Jenny Chang-Claude<sup>15,16</sup> ([j.chang-claude@dkfz-heidelberg.de](mailto:j.chang-claude@dkfz-heidelberg.de)); Anja Rudolph<sup>15</sup> ([a.rudolph@dkfz-heidelberg.de](mailto:a.rudolph@dkfz-heidelberg.de)); Peter Sinn<sup>17</sup> ([Peter.Sinn@med.uni-heidelberg.de](mailto:Peter.Sinn@med.uni-heidelberg.de)); Fergus J Couch<sup>18</sup> ([Couch.Fergus@mayo.edu](mailto:Couch.Fergus@mayo.edu)); Peter Devilee<sup>19</sup> ([p.devilee@lumc.nl](mailto:p.devilee@lumc.nl)); Rob AEM Tollenaar<sup>20</sup> ([R.A.E.M.Tollenaar@lumc.nl](mailto:R.A.E.M.Tollenaar@lumc.nl)); Caroline Seynaeve<sup>21</sup> ([c.seynaeve@erasmusmc.nl](mailto:c.seynaeve@erasmusmc.nl)); Jonine Figueroa<sup>22</sup> ([Jonine.figueroa@ed.ac.uk](mailto:Jonine.figueroa@ed.ac.uk)); Mark E Sherman<sup>23</sup> ([shermans@mail.nih.gov](mailto:shermans@mail.nih.gov)); Jolanta Lissowska<sup>24</sup> ([lissowsj@coi.waw.pl](mailto:lissowsj@coi.waw.pl)); Stephen Hewitt<sup>25</sup> ([hewitts@mail.nih.gov](mailto:hewitts@mail.nih.gov)); Diana Eccles<sup>26</sup> ([D.M.Eccles@soton.ac.uk](mailto:D.M.Eccles@soton.ac.uk)); Maartje J Hooning<sup>27</sup> ([m.hooning@erasmusmc.nl](mailto:m.hooning@erasmusmc.nl)); Antoinette Hollestelle<sup>27</sup> ([a.hollestelle@erasmusmc.nl](mailto:a.hollestelle@erasmusmc.nl)); John WM Martens<sup>27</sup> ([j.martens@erasmusmc.nl](mailto:j.martens@erasmusmc.nl)); Carolien HM van Deurzen<sup>28</sup> ([c.h.m.vandeurzen@erasmusmc.nl](mailto:c.h.m.vandeurzen@erasmusmc.nl)); kConFab Investigators<sup>29</sup> ([heather.thorne@petermac.org](mailto:heather.thorne@petermac.org)); Manjeet K Bolla<sup>30</sup> ([mkh39@medschl.cam.ac.uk](mailto:mkh39@medschl.cam.ac.uk)); Qin Wang<sup>30</sup> ([qw232@medschl.cam.ac.uk](mailto:qw232@medschl.cam.ac.uk)); Michael Jones<sup>1</sup> ([Michael.Jones@icr.ac.uk](mailto:Michael.Jones@icr.ac.uk)); Minouk Schoemaker<sup>1</sup> ([Minouk.Schoemaker@icr.ac.uk](mailto:Minouk.Schoemaker@icr.ac.uk)); Jelle Wesseling<sup>31</sup> ([j.wesseling@nki.nl](mailto:j.wesseling@nki.nl)); Flora E van Leeuwen<sup>32</sup> ([f.v.leeuwen@nki.nl](mailto:f.v.leeuwen@nki.nl)); Laura Van 't Veer<sup>31</sup> ([Laura.Vantveer@ucsf.edu](mailto:Laura.Vantveer@ucsf.edu)); Douglas Easton<sup>4,30</sup> ([dfе20@medschl.cam.ac.uk](mailto:dfе20@medschl.cam.ac.uk)); Anthony J Swerdlow<sup>1,33</sup> ([Anthony.Swerdlow@icr.ac.uk](mailto:Anthony.Swerdlow@icr.ac.uk)); Mitch Dowsett<sup>2,34</sup> ([Mitchell.Dowsett@icr.ac.uk](mailto:Mitchell.Dowsett@icr.ac.uk)); Paul D Pharoah<sup>4,30</sup> ([pp10001@medschl.cam.ac.uk](mailto:pp10001@medschl.cam.ac.uk)); Marjanka K Schmidt<sup>31,32</sup> ([mk.schmidt@nki.nl](mailto:mk.schmidt@nki.nl)) and Montserrat Garcia-Closas<sup>23</sup> ([montserrat.garcia-closas@nih.gov](mailto:montserrat.garcia-closas@nih.gov)).

## Author affiliations

<sup>1</sup> Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

<sup>2</sup> Breast Cancer Now Toby Robins Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

<sup>3</sup> Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

<sup>4</sup> Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

- <sup>5</sup> Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
- <sup>6</sup> Centro de Investigacion en Red de Enfermedades Raras (CIBERER), Valencia, Spain
- <sup>7</sup> Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia
- <sup>8</sup> Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global health, The University of Melbourne, Melbourne, Australia
- <sup>9</sup> Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>10</sup> Division of Preventive Oncology, German Cancer Research Center (DKFZ), and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- <sup>11</sup> German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>12</sup> Saarland Cancer Registry, Saarland, Germany
- <sup>13</sup> School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland
- <sup>14</sup> Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland
- <sup>15</sup> Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>16</sup> University Cancer Center Hamburg , University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- <sup>17</sup> Department of Pathology, Institute of Pathology, Heidelberg University Hospital, Germany
- <sup>18</sup> Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
- <sup>19</sup> Department of Human Genetics & Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands
- <sup>20</sup> Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands
- <sup>21</sup> Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
- <sup>22</sup> Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Scotland, UK
- <sup>23</sup> Divisions of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA
- <sup>24</sup> Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland
- <sup>25</sup> Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Rockville, MD, USA
- <sup>26</sup> Faculty of Medicine Academic Unit of Cancer Sciences, Southampton General Hospital, Southampton, UK

<sup>27</sup> Family Cancer Clinic, Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

<sup>28</sup> Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

<sup>29</sup> Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia

<sup>30</sup> Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

<sup>31</sup> Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

<sup>32</sup> Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

<sup>33</sup> Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

<sup>34</sup> Academic Department of Biochemistry, Royal Marsden Hospital, Fulham Road, London

**Corresponding author**

Mustapha Abubakar

Division of Genetics and Epidemiology

Institute of Cancer Research

15 Cotswold Road, Sutton, London, SM2 5NG

United Kingdom

Email: [Mustapha.Abubakar@icr.ac.uk](mailto:Mustapha.Abubakar@icr.ac.uk)

## Abstract

**Background:** The value of KI67 in breast cancer prognostication has been questioned due to concerns on the analytical validity of visual KI67 assessment and methodological limitations of published studies. Here, we investigate the prognostic value of automated KI67 scoring in a large, multi-centre study, and compare this with pathologists' visual scores available in a subset of patients.

**Methods:** We utilised 143 tissue microarrays (TMAs) containing 15,313 tumour tissue cores from 8,088 breast cancer patients from 10 collaborating studies. A total of 1,401 deaths occurred during a median follow-up of 7.5 years. Centralised KI67 assessment was performed using an automated scoring protocol. The relationship of KI67 levels with 10-year breast cancer specific survival (BCSS) was investigated using Kaplan-Meier survival curves and in Cox proportional hazard regression models adjusted for known prognostic factors.

**Results:** Patients in the highest quartile of KI67 (>12% positive KI67 cells) had a worse 10-year BCSS than patients in the lower three quartiles. This association was statistically significant for ER positive [Hazard ratio (HR) at baseline = 1.96 (1.31 – 2.93); P= 0.001] but not for ER negative [1.23 (0.86 – 1.77); P= 0.248] patients (P-heterogeneity= 0.064). In spite of differences in characteristics of the study populations, the estimates of HR were consistent across all studies (P-heterogeneity =0.941 for ER positive and 0.866 for ER negative). Among ER positive cancers, KI67 was associated with worse prognosis in both node negative [2.47 (1.16 – 5.27)] and positive [1.74 (1.05 – 2.86)] tumours (P-heterogeneity = 0.671). Further classification according to ER, PR and HER2 showed statistically significant associations with prognosis among hormone receptor positive, regardless of HER2 status (P-heterogeneity= 0.270) and among triple negative [1.70 (1.02 – 2.84)] patients. Model fit parameters were similar for visual and automated measures of KI67 in a subset of 2,440 patients with information from both sources.

**Conclusion:** Findings from this large-scale multicentre analysis with centrally generated automated KI67 scores show strong evidence in support of a prognostic value for automated KI67 scoring in breast cancer. Given the advantages of automated scoring in terms of its potential for standardisation, reproducibility, and throughput, automated methods appear as promising alternatives to visual scoring for KI67 assessment.

**Key words:** breast cancer, automated KI67, visual KI67, prognostication

## Introduction

Despite endorsements by several international guidelines (1, 2) KI67 is yet to gain widespread application as a prognostic and/or predictive marker in breast cancer (3). This is due, largely, to methodological variability in KI67 scoring (such as antibody type, specimen type, type of fixative, antigen retrieval methods, method of scoring etc.), and limitations in the design and analyses of studies that have reported on this marker (3-7).

In the majority of settings, KI67 is evaluated visually by a pathologist even though there is yet to be consensus regarding which regions to score between the invasive edge, hot spots or the entire spectrum of the whole section or tumour core (8). As a result, both the intra-observer and, especially, the inter-observer reproducibility of visually derived KI67 scores have been shown to be poor (9-11). This has not only hampered inter-study comparability for KI67, it has fuelled concerns regarding its analytical validity (3). To address some of the methodological issues related to KI67 assessment, the “International KI67 in Breast Cancer Working Group” published recommendations aimed at the standardisation of the analytical processes for KI67 evaluation (8). This panel however fell short of making recommendations regarding the preferred method of scoring for KI67 between visual and automated. Several reports suggest that automated methods could address some of the problems associated with visual scoring (11-19). These methods are high-throughput and not limited by intra-observer variability. However, concerns exist regarding the accuracy of automated methods and the prognostic power of KI67 derived using these methods relative to that derived visually by pathologists. Few relatively small studies have reported a head-to-head comparison between scores derived using both methods in terms of prognostic properties and the results from these are conflicting (11, 17-19).

The majority opinion regarding the prognostic property of KI67 derives mostly from reviews and meta-analyses, which support its prognostic role in breast cancer (4-7, 20). The meta-analyses by de Azambuja et al (6) involving 12,155 patients and another by Harris et al (7) which included over 15,000 patients represent two comprehensive analyses on this subject. These are however limited by

reported evidence of publication bias, significant between-study heterogeneity, and by the fact that most of the included studies utilised different methodological approaches to KI67 evaluation. Furthermore, while the analysis by de Azambuja and colleagues (6) was limited by its inclusion of only univariate hazard ratios, that by Harris and colleagues (7) was limited by the small intersection between the sets of covariates in the included studies. In a population-based cohort of a cancer registry, Inwald and colleagues (21) examined the prognostic role of KI67 in 3,658 patients for whom KI67 was routinely measured in clinical practice and reported significant associations between KI67 and overall survival (21). An important strength of this analysis was that it utilised routinely assessed KI67 measurements in a clinical setting. But this was also limited by the heterogeneity of the KI67 analytical processes in the different laboratories involved in the study. Nonetheless, KI67 has found use in a variety of clinical and epidemiological scenarios including its endorsement by a number of international guidelines for use in treatment decision making in ER positive breast cancer (1, 2) and its incorporation as part of emerging prognostic tools such as the IHC4 score (22, 23) and PREDICT – a breast cancer treatment benefit tool (24).

In this study, we evaluate the value and robustness of automated scoring of KI67 for large-scale, multicentre studies of breast cancer prognostication. We centrally generated automated KI67 score from stained tissue microarrays (TMAs), and assessed its prognostic value overall for different subtypes of breast cancer. We also compared the prognostic performance of automated and visually derived KI67 scores in a subset of patients.

## **Materials and Methods**

### **Study population and study design**

This analysis was conducted within the Breast Cancer Association Consortium which is a large, ongoing collaborative project involving study groups across the globe (25). **Figure 1** shows that we collected a total of 166 TMAs containing 19,039 cores, representing 10,005 patients from 13 study groups (**supplementary Table 1**). Ten study groups provided unstained TMAs which were then

stained and digitised in the Breakthrough Core Pathology laboratory at the Institute of Cancer Research (ICR) and the academic biochemistry laboratory of the Royal Marsden Hospital (RMH), London. Two groups (MARIE and PBCS) provided pre-stained TMAs which were also digitised in our centre. One study (SEARCH) provided TMA images acquired using a similar Ariol technology (a digital image acquisition and analysis system) to the one adopted for this analysis. Of the 10,005 patients, 1,917 were excluded on account of failing predefined quality control checks (N = 946) or due to absent data on follow-up times and/or vital status (N = 971). As a result, a total of 8,088 patients from 10 study groups with a median follow-up of 7.5 years and a total of 1,401 breast cancer specific deaths were used in the survival analysis involving automated KI67. Of these, 2,440 patients with pathologists' visual in addition to automated KI67 scores were used to extrapolate a visual from an automated cut-off point, following which comparative survival analyses involving visual and automated KI67 scores were conducted. Information on other clinico-pathological characteristics of tumours including histological grade, nodal status, tumour size, stage, adjuvant systemic therapy (endocrine therapy and/or chemotherapy) and other IHC markers i.e. oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) were obtained from clinical records. Additional Ariol HER2 data were obtained for a subset of patients with missing clinical HER2 data but for whom data on ER and PR were available (N = 403). All patients provided written informed consent and all participating studies gained approval from the local Ethical Committees and Institutional Review Boards.

### **KI67 immunostaining**

Sections were dewaxed using xylene and rehydrated through graded alcohol (100, 90, and 70%) to water. Slides were then placed in a preheated (5 mins 800W microwave) solution of Dako Target Retrieval solution pH6.0 (S1699) and microwaved on high power for 10mins and then allowed to cool in this solution at room temperature for ten minutes. In the next stage, the slides were placed on a Dako autostainer and stained using a standard protocol using Dako MIB-1 diluted 1/50 and visualised using the Dako REAL kit (K5001). The MIB-1 antibody was also adopted for the staining of those

TMA's that were not part of those centrally stained at the ICR but at varying concentrations (PBCS = 1:500; MARIE = 1:400 and SEARCH = 1:200) (**supplementary Table 2**).

### **KI67 scoring**

This has been previously described (26), but briefly: all TMA's were digitised using the Ariol 50s digital scanner. Fifteen TMA's were selected as a training set. These were scored visually by a pathologist (MA) using a computer-assisted visual (CAV) counting method and used to validate the automated method. The CAV method relied upon built-in features of the Ariol digital system to count negative and positive nuclear populations within 250µm by 250µm squares separated by grids. The standard CAV approach entailed the counting of at least 1,000 cells across the entire spectrum of each core. In the majority of cores, more than 1,000 cells were counted even though fewer than this number was counted in a small minority. Overall, cores with more than 500 cells were considered to be of satisfactory quality. The CAV method is precise, prevents double counting and was observed to have excellent intra-observer reproducibility when a random subset of cores (N = 111) were re-scored at an interval of three months from the first time they were scored (observed agreement/kappa = 96%/0.90); good core-level agreements with two other independent scorers (observed agreement/kappa: CAV vs scorer 2 = 87%/0.66; CAV vs scorer 3 = 84%/0.59; scorer 2 vs. scorer 3 = 89%/0.69) were also observed in a randomly selected subset of 202 cores. Visual scoring in the external TMA's involved both quantitative and semi-quantitative methods. Each core from each patient was scored by two independent pathologists and the KI67 score for each patient was then taken as the average score from the two scorers across all the cores for that patient.

The automated scoring was performed using the Ariol machine which has functionality that allows for the automatic detection of malignant and non-malignant nuclei using shape and size characteristics. Using colour deconvolution, it also distinguishes between DAB positive and negative (haematoxylin) malignant cells. To determine the negative and positive populations of cells, an appropriate region of interest of the malignant cell population in a core was demarcated and two colours were selected to indicate positive and negative nuclear populations. The appropriate colour pixels were then selected to



represent the full range of hue, saturation and intensity that was considered representative of the positive and negative nuclear classes (26). Subsequently, the best shape parameters that discriminated malignant and non-malignant cells according to their spot width, width, roundness, compactness and axis ratio were then also selected. The data were divided into a training and a validation subset and the automated and visual scoring for KI67 showed good agreement (observed agreement = 87%; Kappa = 0.64) and discriminatory accuracy (AUC = 85%) in the validation subset hence allowing for the adoption of this method for the scoring of all 166 TMA's.

### **Statistical methods**

For patients with multiple cores from the same tumour, we used the average KI67 score across valid cores to represent the % positive cells in that tumour. Descriptive analyses of the distribution of KI67 according to clinical and pathological characteristics of the patients were conducted using the non-parametric Kruskal-Wallis equality of medians test for continuous measures and the paired chi-squared test for categorical measures. The relative survival probabilities for patients in different quartiles of the KI67 distribution were compared using Kaplan-Meier survival curves for the 10-year breast cancer specific survival (BCSS). To allow for prevalent cancers, time at risk was left-censored for study entry. It was decided, a priori, to not make any assumptions on a prognostic cut-off point for automated KI67 scores in our dataset but to, instead, leverage on the continuous values to observe a prognostic threshold. As a result, we performed quartile analysis by dividing the continuous KI67 scores into quartiles (Q1-Q4) and examining the prognostic differences among the different quartiles for all the patients in the study. The 10-year BCSS was determined in Kaplan-Meier survival curves and in Cox-proportional hazards regression models stratified by ER status (positive vs negative) and according to nodal status (positive vs negative) and other IHC markers. The univariate Cox models were partially adjusted for study group and age at diagnosis while the multivariate models had further adjustments for other known prognostic factors including histological grade, tumour size, nodal status, morphology, ER, PR, HER2 and adjuvant systemic therapy (endocrine and/or chemotherapy). In the multivariate models, missing values on other covariates were addressed using the multiple imputation

plus outcome (MI+) approach (27). Due to observed violation of the proportionality assumption of the Cox model by automated KI67, it was modelled as a time varying covariate using an extension of the Cox model that allows for the inclusion of a coefficient (T) that varied as an exponential function of time. The log of the coefficient is indicative of both the direction and the magnitude of change in hazard ratio with time, such that if  $\log T$  is  $<1$  hazard falls with time while if it is  $>1$  hazard increases with time. Known violation of the proportional hazards assumption by ER was addressed in the same way. Consistency of hazard ratio (HR) estimates across the different study groups was evaluated using the  $I^2$  statistic, derived by performing a fixed-effect meta-analysis of study-specific HR estimates. To enable direct comparison between the visual and automated KI67 scores, we extrapolated a visual from an automated cut-off point in a linear regression model and used the resulting cut-off point for all further analyses. All analyses were conducted using STATA statistical software version 10 (StataCorp, 4905 Lakeway drive, College Station, Texas). Statistical tests were two-sided and p-values less than 0.05 were considered to be statistically significant.

## **Results**

### **Description of study population and association between automated KI67 score and other clinico-pathological characteristics of breast cancer patients (N = 8,088)**

In all, a total of 143 TMAs containing 15,313 cores from 8,088 patients were used in this analysis as shown in **Figure 1**. The studies included in this analysis used different TMA designs (**Table 1**). More than half (4,431/55%) of the patients had KI67 scores on at least two cores and evaluation of dichotomous categories revealed concordant KI67 status in 83.7% of the patients. When we examined the distribution of continuous KI67 scores among categories of the different clinical and pathological characteristics we observed this to differ according to histological grade, tumour size, morphology, ER status, PR status, and HER2 status but not nodal status or stage at diagnosis (**Figure 2**). The distribution of these characteristics for patients with high (Q4 or  $>12\%$  positive cells) and low (Q1-Q3) KI67 are shown in **Supplementary Tables 3 and 4** for ER positive and ER negative patients, respectively.

### Association between automated KI67 score and 10-year BCSS among 8,088 patients

Using continuous measures of KI67 categorised into quartiles, we observed poorest survival in the highest quartile, corresponding to 12% positive cells, but little difference in survival between the other three (Q1-Q3) quartiles (log-rank  $P = 1.2 \times 10^{-5}$ ; **Figure 3A**). As a result, the continuous KI67 value was dichotomised at the threshold of 12% in subsequent analyses. High KI67 was significantly associated with 10-year BCSS overall (log-rank  $P = 3.1 \times 10^{-7}$ ) among ER positive (log-rank  $p$ -value =  $1.3 \times 10^{-3}$ ) but not ER negative (log-rank  $P = 0.35$ ) cancers (**Figure 3**, B, C and D, respectively). Similarly, in multivariate models, high KI67 expression was significantly associated with worse 10-year BCSS among ER positive (HR at baseline = 1.96; 95% CI = 1.31 – 2.93) but not ER negative breast cancers (HR = 1.23; 95% CI = 0.86 – 1.77; P-heterogeneity = 0.064) (**Table 2**). Further stratification of ER positive cancers according to nodal status showed that high KI67 was associated with worse survival in both node negative and node positive cancers in multivariate analysis [Node negative (2.47 (1.16 – 5.27)); Node positive (1.74 (1.05 – 2.86)); P-heterogeneity = 0.67] (**Table 2**). The association between KI67 and survival was significant among ER positive patients who did not receive chemotherapy [1.95 (1.18 – 3.21);  $p$ -value = 0.009] but not among those who did [1.89 (0.84 – 4.29);  $p$ -value = 0.124; P-heterogeneity = 0.60]. We found no evidence of between study heterogeneity in estimates of HR for ER positive ( $I^2 = 0.0\%$ ,  $P = 0.94$ ) or ER negative ( $I^2 = 0.0\%$ ,  $P = 0.86$ ) patients (**supplementary Figure 1**). Among hormone receptor positive breast cancers, the HR for KI67 was not significantly different according to HER2 status (**Table 2**; P-heterogeneity = 0.270). Modest evidence for a poorer prognosis among high, relative to low, KI67 was also seen for triple negative breast cancers [1.70 (1.02 – 2.84);  $P = 0.04$ ]. No significant associations with prognosis were found for KI67 among HER2-positive (i.e. ER-/PR-/HER2+) [0.91 (0.60 – 1.36)] breast cancers (**Table 2**).

### **Comparison of 10-year BCSS among 2,440 patients with both visual and automated quantitative KI67 scores**

The automated cut-off point of 12% positive cells corresponded to a visual cut-off point of 24.2% based on a linear regression model comprising patients with quantitative data on both methods. The visual cut-off was rounded up to a cut-off point of 25%. Strong evidence ( $p$ -value  $< 0.0001$ ) in support of a positive linear correlation ( $r = 0.63$ ) between automated and visual scores was observed and continuous automated scores showed good discriminatory accuracy against the visually determined binary classes (AUC = 82%, 95%CI = 80 – 84%). 26% of the patients were classified as having high visual KI67 in contrast to 29% for the automated KI67 scores; cross classification of visual and automated categories revealed better specificity (84%) than sensitivity (65.6%) for the automated score in classifying visually determined categories (**Supplementary Table 5**). High KI67 was associated with worse survival in Kaplan-Meier curves based on both automated (log-rank  $p$ -value =  $9.8 \times 10^{-6}$ ) and visual (log-rank  $p$ -value =  $3.8 \times 10^{-14}$ ) KI67 even though attenuation of the difference between strata was observed for automated KI67 scores (**Supplementary Figure 3**). In two separate models for visual and automated KI67 each adjusted for age at diagnosis and study group we observed stronger evidence for an association between KI67 and survival for visual than automated KI67 (Table 3). Analysis of model fit however revealed similar parameters for both scores especially in ER positive breast cancers (AIC/BIC: visual = 2656/2618 and automated 2675/2638) (**Table 3**). When we performed further adjustments for other prognostic factors in multivariate Cox models of imputed datasets, we observed both visual and automated KI67 scores to be significantly associated with survival for all patients [HR (95%CI): visual = 1.75 (1.23 – 2.49); automated = 1.61 (1.14 – 2.28)] and in ER positive [visual = 2.30 (1.34 – 3.94); automated = 2.10 (1.28 – 3.47)] but not ER negative [visual = 1.63 (0.97 – 2.72); automated = 1.28 (0.79 – 2.05)] patients (**Table 3**).

## Discussion

Findings from our analysis provide strong evidence in support of a prognostic relationship for automated KI67 scoring in ER positive (node negative and node positive) patients that is independent of tumour grade and other prognostic factors. Even though our data suggested a larger magnitude of the association between KI67 and survival among the node negative patients, the difference between node positive and negative was not statistically significant. With over 8,000 patients from multiple centres internationally, this represents the largest study that has evaluated the prognostic value of automated KI67 scoring in breast cancer till date. Furthermore, the large sample size allowed us to evaluate its prognostic value in a number of breast cancer subtypes including ER+ (node negative and positive), ER-, ER+ and/or PR+ (HER2+ or HER2-), ER-/PR- and HER2+ (i.e. HER2-enriched) and triple negative breast cancers.

Our findings suggest that automated KI67 scoring is an analytically valid approach to generating KI67 scores. This is particularly noteworthy given the growing need to incorporate measures of KI67 in prognostic tools such as the IHC4 score and PREDICT (23, 24). These tools are relatively cheap, readily available, and utilise routinely measured IHC markers and, in the case of PREDICT, other routinely available patient data to provide information that can help clinicians and patients make informed decisions regarding the course of treatment. It is acknowledged that prognostication in breast cancer is becoming increasingly more sophisticated and that a number of multigene assays (28, 29) have been validated for this purpose, however, their costs and proprietary concerns limit their use in a large number of settings. Moreover, findings from previous studies suggest that some multigene assays may not perform better than routinely measured IHC markers. For instance, Cuzick et al (23) reported similar prognostic properties for the Genomic Health recurrence score (GHI-RS, Oncotype DX), a 21 gene panel test, and the IHC4 score in their analysis of 1,125 women from the TransATAC study and notably KI67 was assessed by image analysis in that study (23). Nonetheless, the relative performance of visual and automated KI67 scores in relation to the IHC4 score or PREDICT can only be assessed in studies that are specifically designed for that purpose.

In addition to lack of analytical validity, the prognostic performance of KI67 has also been questioned due to the design and analysis of studies that have previously reported on this protein (3). Ours is a large scale, multicentre analysis which has adopted the recommended laboratory processes for the staining and scoring of KI67 (8). All TMAs in our analysis were stained using the MIB1 antibody (even though not all of them were centrally stained in our centre) and scored using a single automated algorithm. Our estimates of a ~2-fold and ~1.5-fold increased risk of mortality at baseline for high versus low KI67 in univariate and multivariate analyses, respectively, are similar to those reported by de Azambuja et al (HR = 1.95) and Harris et al (HR = 1.42) in their univariate and multivariate meta-analyses, respectively (6, 7). Stratification of our analysis according to other IHC markers (in addition to ER) showed automated KI67 to be prognostic in hormone receptor positive cancers. These findings, together with our observation of the prognostic value of KI67 in both node negative and node positive ER positive patients, support the decision by the St Gallen international Expert Consensus to endorse KI67 for treatment decision making in ER positive early (1-3 axillary nodes) breast cancer patients (1). We also observed modest evidence in support of poorer survival outcomes among high, relative to low, KI67 expressing triple negative subtypes of breast cancer. This finding is in support of a previous report by Keam and colleagues (30). Our population of triple negative breast cancers (N=1,001) was however 9.5 times larger than that of Keam et al (N = 105).

Comparative analysis of visual and automated KI67 scores showed a stronger survival association for the visual over the automated scores; however, differences were generally modest. Given the advantages of automated versus visual scoring in terms of its potential for standardisation, reproducibility, and throughput, automated methods appear as promising alternatives to visual scoring for KI67 assessment. A potential limitation to the adoption of automated KI67 scoring in the clinical setting is that misclassification of positive as negative nuclei, or malignant as benign nuclei could lead to attenuation of prognostic associations, an observation that has been previously reported for ER and PR (31) and one which we have also observed for KI67 in this analysis. This can however be mitigated by stringent quality control processes or by the adoption of a synergistic approach that combines the benefits of both the automated and visual scoring methods. One such approach is the

CAV scoring method which we developed for the visual counting of negative and positive malignant nuclei. This approach, a variation of which has been previously reported (15), exploits the advantages of both visual and digital imaging tools by enabling the visual counting of KI67 positive cells in well-defined areas of a tumour within a computer microenvironment. This method is however limited by the observation that it is time consuming; as such, it may not be efficient if adopted for the large-scale scoring of KI67 in epidemiological studies, clinical trials or biomarker discovery studies. Nonetheless, efforts are currently underway to standardise the methods for the visual scoring of KI67 in core-cuts.

We centrally generated KI67 scores on TMAs and determined a threshold of 12% positive cells of prognostic relevance in our study population. However, due to possible variations in the distribution of KI67 scores according to specimen type and among different laboratories, this cut-off point may not apply to other types of clinical samples or to other laboratories. As a result, pending international standardisation of the KI67 analytical processes, setting local laboratory-specific cut-off points as recommended by international guidelines (1) remains a pragmatic approach to determining 'high' and 'low' KI67. Furthermore, although our automated cut-off point of 12% positive cells was determined to correspond to a visual score of 25%, this may be related, at least in part, to the fact that automated systems generally count more cells than the visual evaluator, a reason that has been proposed to explain differences in KI67 scores between visual and automated scoring and between different automated scoring approaches (26). Nonetheless, findings from a recent meta-analysis that assessed the prognostic value of different cut-off levels of KI67 suggest that a visual cut-off point of >25% provides greater discrimination in mortality risk than other cut-off points (32).

Some limitations of our analysis include the lack of data on specific chemotherapeutic or endocrine agents received by each patient, as a result of which we were unable to account for the impact of specific treatment regimen on survival or to examine whether or not KI67 is predictive of response to specific chemotherapeutic and/or endocrine agents. We were however able to account for whether or not patients received adjuvant systemic treatment in all our analyses as more than two-thirds of the patients had information on treatment. This also allowed us to perform stratified analysis according to

whether or not chemotherapy was administered. Also, we did not have data on disease free survival which may have been a more informative end point than BCSS in early breast cancer. Our assessment of KI67 on TMAs may mean that direct inference cannot be drawn from our findings on other types of clinical samples, especially whole sections (8). This is because KI67 scores are speculated to be lower for TMAs than for whole sections and not many studies have assessed the correlation between KI67 scores on TMAs and those on whole sections. However, one such study by Kobierzycki and colleagues involving 51 archival paraffin blocks of invasive ductal carcinoma showed excellent correlation ( $r = 0.91$ ) between the TMAs and whole sections (33). Their paper however utilised three 0.6 mm core punches and this may explain the high correlation between KI67 scores on TMAs and whole sections that was observed in that study. Nonetheless, the fact that more than half (4,431/55%) of the patients in our analysis had KI67 scores on two or more cores with 83% of these showing concordant KI67 status should limit the impact of intra-tumour heterogeneity of KI67 scores on our findings.

## **Conclusion**

Our large, multicentre study indicates that automated KI67 scoring provides prognostic information in breast cancer that is independent of standard parameters. In view of its potential for standardisation, throughput and reproducibility, the automated method appears to be a promising alternative to visual scoring for KI67. These findings are important given the increasing need to incorporate measures of KI67 as part of tools that are needed to refine prognostic scores for breast cancer patients; this is especially relevant for patients with ER-positive, node negative tumours, in order to aid decisions on providing adjuvant chemotherapy. However, further work is needed to standardise the staining and scoring protocols for KI67. In doing so, the potential benefits and drawbacks of automated versus visual scoring systems should merit consideration. It is in light of this that we welcome ongoing efforts by the International Working Party on KI67 in Breast cancer aimed at standardisation of the analytical processes for KI67.

## **Declarations**



## **Abbreviations**

TMA, Tissue microarray; BCAC, Breast Cancer Association Consortium; ICR, Institute of Cancer Research; DAB, Diaminobenzidine; ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; CAV, Computer assisted visual; AUC, area under the curve; ROC, receiver operating characteristic; MI, multiple imputation; BCSS, breast cancer specific survival; HR, hazard ratio; GHI-RS, Genomic Health recurrence score; ICC, intra-class correlation; IHC, immunohistochemical.

## **Ethical approval and consent to participate**

Each of the individual studies was approved by the local Ethics Committees and Institutional Review Boards (IRBs) and written informed consent to participate in the study was obtained from each participant across all study groups. The ABCS study received approval of the Medical Ethical Committees of the Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital [NKI-AVL; Amsterdam, the Netherlands] and Leiden University Medical Cancer [LUMC; Leiden, the Netherlands]. The ESTHER study was approved by the Ethics Committees of the Medical Faculty of the University of Heidelberg and the Medical Association of Saarland. The joint Ethics Committee of Kuopio University and Kuopio University Hospital approved the Kuopio Breast Cancer Project (KBCP). Approval for the MARIE study was obtained from the Ethics Committee of the University of Heidelberg, the Hamburg Medical Council, and the Medical Board of the State of Rheinland-Pfalz. MCBCS study was approved by the Ethics Committee of the Mayo Clinic College of Medicine. The Medical Ethical Review Boards of the Rotterdam Cancer Centre and academic cancer centre in Leiden approved the study protocol for ORIGO study. The PBCS study protocol was reviewed and approved by local and the US National Cancer Institute (NCI) IRBs. The RBCS study was approved by the Ethical Committees of the University Hospital Rotterdam, Erasmus University Rotterdam and Leiden University Medical Centre, Leiden. The SEARCH study was approved by the Eastern multi-centre research ethics committee. The kConFab study obtained human ethics approval at all the participating institutions through which subjects are recruited.

## **Consent for publication**

Not applicable

## **Availability of supporting data**

The data that support the findings of this study belong to the Breast Cancer Association Consortium (BCAC) and cannot be shared in a public repository as participants have not consented to sharing their data publicly.

### **Competing interests**

The authors declare that they have no competing interests.

### **Funding**

ABCS was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]; BBMRI-NL, which is a Research Infrastructure financed by the Dutch government (NWO 184.021.007); and the Dutch National Genomics Initiative.

The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe).

The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland.

We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia.

The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].

The MCBCS was supported by an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], the Breast Cancer Research Foundation, the Mayo Clinic Breast Cancer Registry and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation.

ORIGO authors thank E. Krol-Warmerdam, and J. Blom; The contributing studies were funded by grants from the Dutch Cancer Society (UL1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16).

PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318).

SEARCH is funded by programme grant from Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. Part of this work was supported by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009223175) (COGS). We acknowledge funds from Breakthrough Breast Cancer, UK, in support of MGC at the time this work was carried out and funds from the Cancer Research, UK, in support of MA at the Institute of Cancer Research, London.

### **Authors' contributions**

MA and MG-C conceived and carried out the analysis; MG-C supervised the work; FD carried out centralised laboratory work. MA developed the automated and CAV KI67 scoring protocols; NO, AJS and MD provided additional supervisory support; PC performed data management; MA and MG-C analysed the data; FB, HRA, PC, JB, RM, HB, CS, AM, JCC, AR, PS, FJC, PD, JF, MES, JL, DE, MJH, AH, JWMM, CHMvD, MKB, QW, MJ, MS, AJS, DE, JW, LV'tV, FEvL, MKS, PDP contributed to TMA/data collection and/or data management. All authors contributed to manuscript development and writing and gave final approval for its submission.

### **Acknowledgements**

We acknowledge support from Will Howat and Leigh-Anne McDuffus of the Cancer Research UK Cambridge Institute, University of Cambridge and Lila Zabaglo of the Academic Biochemistry Laboratory of the Institute of Cancer research, London during the development of the automated scoring protocol.

### **Authors' information**

Not applicable

## References

- [1]. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*. 2013 August 4, 2013.
- [2]. Senkus E, Kyriakides S, Penault-Llorca F, Poortmans P, Thompson A, Zackrisson S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2013 August 22, 2013.
- [3]. Andre F, Arnedos M, Goubar A, Ghouadni A, Delaloge S. Ki67[mdash]no evidence for its use in node-positive breast cancer. *Nat Rev Clin Oncol*. 2015 03/17/online;advance online publication.
- [4]. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. *Journal of Clinical Oncology*. 2005;23(28):7212-20.
- [5]. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *The lancet oncology*. 2010;11(2):174-83.
- [6]. de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. *British journal of cancer*. 2007;96(10):1504-13.
- [7]. Stuart-Harris R, Caldas C, Pinder S, Pharoah P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *The Breast*. 2008;17(4):323-34.
- [8]. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *Journal of the National Cancer Institute*. 2011;103(22):1656-64.
- [9]. Polley M-YC, Leung SCY, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An International Ki67 Reproducibility Study. *Journal of the National Cancer Institute*. 2013 December 18, 2013;105(24):1897-906.
- [10]. Mikami Y, Ueno T, Yoshimura K, Tsuda H, Kurosumi M, Masuda S, et al. Interobserver concordance of Ki67 labeling index in breast cancer: Japan Breast Cancer Research Group Ki67 Ring Study. *Cancer Science*. 2013;104(11):1539-43.
- [11]. Gudlaugsson E, Skaland I, Janssen EAM, Smaaland R, Shao Z, Malpica A, et al. Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. *Histopathology*. 2012;61(6):1134-44.
- [12]. Pinder S, Wencyk P, Sibbering D, Bell J, Elston C, Nicholson R, et al. Assessment of the new proliferation marker MIB1 in breast carcinoma using image analysis: associations with other prognostic factors and survival. *British journal of cancer*. 1995;71(1):146.
- [13]. Pietiläinen T, Lipponen P, Aaltomaa S, Eskelinen M, Kosma V, Syrjänen K. The important prognostic value of Ki-67 expression as determined by image analysis in breast cancer. *Journal of cancer research and clinical oncology*. 1996;122(11):687-92.
- [14]. Fasanella S, Leonardi E, Cantaloni C, Eccher C, Bazzanella I, Aldovini D, et al. Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies. *Diagnostic Pathology*. 2011 03/30;6(Suppl 1):S7-S. PubMed PMID: PMC3073225.
- [15]. Laurinavicius A, Plancoulaine B, Laurinaviciene A, Herlin P, Meskauskas R, Baltrusaityte I, et al. A methodology to ensure and improve accuracy of Ki67 labelling index estimation by automated digital image analysis in breast cancer tissue. *Breast Cancer Res*. 2014;16:R35.
- [16]. Tuominen VJ, Ruotoistenmaki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res*. 2010;12(4):R56.
- [17]. Konsti J, Lundin M, Joensuu H, Lehtimäki T, Sihto H, Holli K, et al. Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer. *BMC clinical pathology*. 2011;11(1):3.

- [18]. Mohammed ZMA, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, et al. Comparison of Visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *Br J Cancer*. 2012 01/17/print;106(2):383-8.
- [19]. Klauschen F, Wienert S, Schmitt WD, Loibl S, Gerber B, Blohmer J-U, et al. Standardized Ki67 Diagnostics Using Automated Scoring—Clinical Validation in the GeparTrio Breast Cancer Study. *Clinical Cancer Research*. 2015 August 15, 2015;21(16):3651-7.
- [20]. Sheri A, Dowsett M. Developments in Ki67 and other biomarkers for treatment decision making in breast cancer. *Annals of Oncology*. 2012;23(suppl 10):x219-x27.
- [21]. Inwald EC, Klinkhammer-Schalke M, Hofstädter F, Zeman F, Koller M, Gerstenhauer M, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res Treat*. 2013 2013/06/01;139(2):539-52. English.
- [22]. Cuzick J, Dowsett M, Wale C, Salter J, Quinn E, Zabaglo L, et al., editors. Prognostic Value of a Combined ER, PgR, Ki67, HER2 Immunohistochemical (IHC4) Score and Comparison with the GHI Recurrence Score—Results from TransATAC. *Cancer Research*; 2009: Amer Assoc Cancer Research 615 Chestnut St, 17th Floor, Philadelphia, PA 19106-4404 USA.
- [23]. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, 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. *Journal of Clinical Oncology*. 2011;29(32):4273-8.
- [24]. Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer*. 2012 08/21/print;107(5):800-7.
- [25]. BCAC. Breast Cancer Association Consortium [cited 2015 19/08/2015]. Available from: <http://apps.ccge.medschl.cam.ac.uk/consortia/bcac/index.html>.
- [26]. Abubakar M, Howat WJ, Daley F, Zabaglo L, McDuffus L-A, Blows F, et al. High-throughput automated scoring of Ki67 in breast cancer tissue microarrays from the Breast Cancer Association Consortium. *The Journal of Pathology: Clinical Research*. 2016:n/a-n/a.
- [27]. Ali AMG, Dawson SJ, Blows FM, Provenzano E, Ellis IO, Baglietto L, et al. Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. *Br J Cancer*. 2011 02/15/print;104(4):693-9.
- [28]. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *New England Journal of Medicine*. 2004;351(27):2817-26. PubMed PMID: 15591335.
- [29]. Filipits M, Nielsen TO, Rudas M, Greil R, Stöger H, Jakesz R, et al. The PAM50 Risk-of-Recurrence Score Predicts Risk for Late Distant Recurrence after Endocrine Therapy in Postmenopausal Women with Endocrine-Responsive Early Breast Cancer. *Clinical Cancer Research*. 2014 March 1, 2014;20(5):1298-305.
- [30]. Keam B, Im S-A, Lee K-H, Han S-W, Oh D-Y, Kim JH, et al. Ki-67 can be used for further classification of triple negative breast cancer into two subtypes with different response and prognosis. *Breast Cancer Res*. 2011;13(2):R22.
- [31]. Howat WJ, Blows FM, Provenzano E, Brook MN, Morris L, Gazinska P, et al. Performance of automated scoring of ER, PR, HER2, CK5/6 and EGFR in breast cancer tissue microarrays in the Breast Cancer Association Consortium. *The Clinical Journal of Pathology*. 2014:n/a-n/a.
- [32]. Petrelli F, Viale G, Cabiddu M, Barni S. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. *Breast Cancer Res Treat*. 2015;153(3):477-91.
- [33]. Kobierzycki C, Pula B, Wojnar A, Podhorska-Okolow M, Dziegiel P. Tissue Microarray Technique in Evaluation of Proliferative Activity in Invasive Ductal Breast Cancer. *Anticancer Research*. 2012 March 1, 2012;32(3):773-7.

**Table 1: Description of study populations, TMA designs and patient characteristics for the 8,088 patients included in this analysis**

Study Acronym	Country	Cases (N)	Age at diagnosis	TMAs	Coresh/patient	Coresh/patient	Coresh	Core size	Total coresh
			Mean (range)		(range)	(average)	per TMA	in mm	per study
ABCS	Netherlands	885	43 (19-50)	24	1-6	2.77	15-328	0.6	2,449
ESTHER	Germany	252	62 (50-75)	6	1-2	1.83	78-91	0.6	461
KBCP	Finland	266	59 (30-92)	12	1-3	2.72	63-94	1.0	724
MARIE	Germany	808	62 (50-75)	27	1-5	1.84	32-92	0.6	1,490
MCBCS	USA	437	58 (22-87)	7	1-8	3.73	131-301	0.6	1,630
ORIGO	Netherlands	348	53 (22-87)	9	1-9	2.85	67-223	0.6	991
PBCS	Poland	1,068	56 (27-75)	22	1-2	2.21	66-145	1.0	2,358
RBCS	Netherlands	225	45 (25-84)	6	1-5	2.85	134-199	0.6	642
SEARCH	United Kingdom	3,491	52 (24-70)	24	1-3	1.16	120-167	0.6	4,037
kConFab	Australia/New Zealand	308	45 (20-77)	6	1-2	1.72	65-114	0.6	531
Total		8,088	53 (19-92)	143	1-9	1.89	15-328	0.6-1.0	15,313

**Table 2: Hazard ratios (HR) and 95% CI for the association between automated KI67 score and 10-year BCSS in partially and fully adjusted models: Analysis stratified overall and according to ER, nodal status and other IHC markers (N = 8,088 patients).**

	Patients (N)	Deaths (N)	Partially adjusted <sup>1</sup>		Fully adjusted <sup>2*</sup>	
			HR (95%CI)	p-value	HR (95%CI)	p-value
<b>All cancers</b>						
Low KI67	6,093	1,030	1.00 (referent)		1.00 (referent)	
High KI67	1,995	371	2.32 (1.79, 3.00)	6.34×10 <sup>-11</sup>	1.47 (1.13, 1.92)	0.004
T <sup>3</sup>			0.89 (0.84, 0.94)	3.20×10 <sup>-05</sup>	0.93 (0.87, 0.98)	0.010
<b>ER positive</b>						
Low KI67	4,379	615	1.00 (referent)		1.00 (referent)	
High KI67	1,141	166	2.47 (1.63, 3.72)	8.45×10 <sup>-06</sup>	1.96 (1.31, 2.93)	0.001
T			0.88 (0.81, 0.96)	0.003	0.89 (0.82, 0.97)	0.006
<b>ER negative</b>						
Low KI67	1,271	320	1.00 (referent)		1.00 (referent)	
High KI67	778	188	1.38 (0.97, 1.97)	0.072	1.23 (0.86, 1.77)	0.248
T			0.94 (0.86, 1.02)	0.155	0.94 (0.87, 1.03)	0.199
<b>ER positive/Node positive</b>						
Low KI67	1,550	350	1.00 (referent)		1.00 (referent)	
High KI67	408	94	2.22 (1.31, 3.77)	0.003	1.74 (1.05, 2.86)	0.031
T			0.91 (0.82, 0.99)	0.044	0.91 (0.82, 1.01)	0.075
<b>ER positive/Node negative</b>						
Low KI67	2,399	205	1.00 (referent)		1.00 (referent)	
High KI67	561	55	3.17 (1.43, 6.99)	0.004	2.47 (1.16, 5.27)	0.019
T			0.84 (0.71, 0.98)	0.034	0.86 (0.73, 0.99)	0.048
<b>HRP/HER2-</b>						
Low KI67	3,332	462	1.00 (referent)		1.00 (referent)	
High KI67	831	114	1.69 (1.26, 2.27)	2.42×10 <sup>-04</sup>	1.49 (1.10, 2.00)	0.009
T			0.94 (0.90, 0.98)	0.004	0.94 (0.90, 0.98)	0.004
<b>HRP/HER2+</b>						
Low KI67	421	82	1.00 (referent)		1.00 (referent)	
High KI67	157	36	1.96 (1.28, 3.00)	9.70×10 <sup>-04</sup>	1.59 (1.03, 2.45)	0.035
T			0.94 (0.90, 0.98)	0.004	0.94 (0.90, 0.98)	0.004
<b>HRN/HER2- (Triple negative)</b>						
Low KI67	565	142	1.00 (referent)		1.00 (referent)	
High KI67	436	107	1.75 (1.06, 2.90)	0.028	1.70 (1.02, 2.84)	0.044
T			0.86 (0.75, 0.98)	0.031	0.86 (0.75, 0.99)	0.031
<b>HRN/HER2+ (HER2-enriched)</b>						
Low KI67	227	85	1.00 (referent)		1.00 (referent)	
High KI67	149	48	0.76 (0.37, 1.55)	0.450	0.75 (0.36, 1.57)	0.455
T			1.08 (0.90, 1.29)	0.396	1.07 (0.90, 1.29)	0.435

<sup>1</sup>Partially adjusted models – adjusted for age at diagnosis and study group

<sup>2</sup>Fully adjusted models – further adjustment for histological grade, tumour size, nodal status, morphology, PR, HER2, systemic therapy (endocrine and/or chemotherapy) and (for the model involving all patients) ER status. \*This model was based on 20 imputations performed to address missing values on other covariates

<sup>3</sup>T = log of time varying coefficient (if T <1 hazard falls with time and if T >1 hazard increases with time).

HRP = **H**ormone **R**eceptor **P**ositive

HRN = **H**ormone **R**eceptor **N**egative

**Table 3: Univariate (partially adjusted) and multivariate (fully adjusted) hazard ratios (HR) and 95% CI for the associations between automated and visual KI67 scores with survival in breast cancer (N = 2,440)**

	<sup>1</sup> Partially adjusted model								<sup>2</sup> Fully adjusted model*			
	Visual				Automated				Visual		Automated	
	Patients	Deaths	HR (95%CI)	p-value	Patients	Deaths	HR (95%CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
<b>Overall</b>												
Low KI67	1,804	116	1.00 (Ref)		1,728	125	1.00 (Ref)		1.00 (Ref)		1.00 (Ref)	
High KI67	636	78	2.40 (1.92, 3.01)	2.20×10 <sup>-14</sup>	712	69	1.67 (1.33, 2.10)	1.30×10 <sup>-05</sup>	1.75 (1.23, 2.49)	0.002	1.61 (1.14, 2.28)	0.007
AIC †			5050.4				5087.2					
BIC †			5090.8				5127.6					
<b>ER positive</b>												
Low KI67	1,337	69	1.00 (Ref)		1,241	69	1.00 (Ref)		1.00 (Ref)		1.00 (Ref)	
High KI67	282	27	2.40 (1.72, 3.33)	2.00×10 <sup>-07</sup>	378	27	1.47 (1.05, 2.04)	0.024	2.30 (1.34, 3.94)	0.002	2.10 (1.28, 3.47)	0.004
AIC			2618.8				2638.2					
BIC			2656.8				2675.8					
<b>ER negative</b>												
Low KI67	357	39	1.00 (Ref)		392	48	1.00 (Ref)		1.00 (Ref)		1.00 (Ref)	
High KI67	331	48	1.84 (1.30, 2.62)	6.10×10 <sup>-04</sup>	296	39	1.44 (1.02, 2.04)	0.043	1.62 (0.97, 2.72)	0.066	1.28 (0.79, 2.05)	0.312
AIC			1755.7				1763.4					
BIC			1787.4				1795.1					

<sup>1</sup> Partially adjusted model – adjusted for age at diagnosis and study group only

<sup>2</sup> Fully adjusted model – further adjustment for other prognostic factors including histological grade, tumour size, nodal status, ER, PR, HER2, morphological subtype, and systemic therapy (endocrine and/or chemotherapy). \*This model was based on 20 imputations performed to address missing values on other covariates.

† Model fit parameters: AIC – Akaike information criterion; BIC – Bayesian information criterion.



## Figure legends

### **Figure 1: Study population and study design**

We collected 166 TMAs containing 19,039 cores from 10,005 patients. Of these, 15 TMAs containing 1,346 cores were selected as the training set and these were used to develop an automated scoring protocol that was validated against corresponding computer-assisted visual (CAV) scores. Ultimately, this protocol was applied to the scoring of all 166 TMAs. Following automated scoring, all cores that failed our priori defined quality control checks (including total nuclei count  $>50$  and  $<15,000$ , and KI67 score = 100%) were excluded (N = 946 patients). For the purpose of survival analyses, all subjects with missing follow-up/survival data were also excluded (N = 971 patients). As a result, a total of 8,088 patients were used in the survival analysis involving automated KI67 score. Furthermore, based on a subset of patients (N = 2,440) with pathologists' in addition to the automated KI67 scores, we extrapolated a visual from an automated cut-off point and used this to compare the prognostic performance of visual and automated KI67 scores in breast cancer.

**Figure 2: Distribution of continuous KI67 scores according to categories of other clinical and pathological variables.** Significant differences were seen in the distribution of automated KI67 scores according to categories of histological grade, tumour size, morphology, ER status, PR status, and HER2 status but not nodal status or stage.

**Figure 3: Kaplan-Meier survival curves for the 10-year BCSS according to strata of automated KI67 scores, overall and by ER status.**

KM survival curves for the association between KI67 and 10-year BCSS among (A) Quartiles of KI67 (Q1 =  $<25^{\text{th}}$  percentile; Q2 =  $25^{\text{th}} - 50^{\text{th}}$  percentile; Q3 =  $>50^{\text{th}} - 75^{\text{th}}$  percentile and Q4 =  $>75^{\text{th}}$  percentile; N = 8,088); (B) Dichotomous categories of KI67 ( $\leq 12\%$ /low and  $>12\%$ /high), overall (N = 8,088 patients); (C) ER positive cancers (N = 5,520 patients); and (D) ER negative cancers (N = 2,049 patients).

## Supplementary Materials

**Supplementary Figure 1:** Meta-analysis of study specific hazard ratios (HR) stratified by ER status

**Supplementary Figure 2:** ROC curve for the discriminatory accuracy of continuous automated KI67 scores against binary visual categories ( $\leq 25\%$  and  $> 25\%$ ) based on a subset of patients with data on both visual and automated KI67 scores.

**Supplementary Figure 3:** Kaplan-Meier survival curves for the 10-year BCSS according to strata (high and low) of automated and visual KI67 scores (N = 2,440)

**Supplementary Table 1:** Description of study populations

**Supplementary Table 2:** KI67 immunohistochemistry reagents and antigen retrieval protocols according to study groups

**Supplementary Table 3:** Association of clinical and pathological characteristics with high ( $>12\%$ ) and low ( $\leq 12\%$ ) KI67 categories among 5,520 ER positive breast cancer cases

**Supplementary Table 4:** Association of clinical and pathological characteristics with high ( $>12\%$ ) and low ( $\leq 12\%$ ) KI67 categories among 2,049 ER negative breast cancer cases

**Supplementary Table 5:** Cross classification of visual and automated KI67 score categories

**Supplementary Table 6:** Multivariate model for the association of KI67 with 10-year BCSS among 5,520 ER positive patients

**Supplementary Table 7:** Multivariate model for the association of KI67 with 10-year BCSS among 2,049 ER negative patients