



# Digital quantification of Ki-67 in breast cancer

María del Rosario Taco Sanchez<sup>1</sup> · Teresa Soler-Monsó<sup>1,2</sup> · Anna Petit<sup>1</sup> · Juan Azcarate<sup>1</sup> · Alba Lasheras<sup>1</sup> · Carmen Artal<sup>1</sup> · Miguel Gil<sup>2</sup> · Catalina Faló<sup>2</sup> · María Jesús Pla<sup>3</sup> · Xavier Matias-Guiu<sup>4</sup>

Received: 8 March 2018 / Revised: 11 October 2018 / Accepted: 23 October 2018 / Published online: 21 November 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Ki-67 proliferative index (Ki-67) is a predictive and prognostic factor in breast cancer (BC). However, some international committees do not recommend its use in routine practice due to insufficient clinical evidence and lack of standardisation and assessment method reproducibility. Scoring of Ki-67 by digital pathology may contribute to overcome these drawbacks. We evaluated 136 core biopsies of BC patients and calculated the correlation of Ki-67 scored by two breast pathologists with two methods, eyeballing visual assessment (EB) on the microscope and digital image analysis (DI), both assessed from hot spot areas (HS) and the average between hot and cold spot areas (AVE). Good and higher correlation between pathologists was observed for HS using DI in comparison to EB (0.861 vs. 0.828). Correlation in HS with both methods was very similar in homogeneous tumours (0.869 vs. 0.866). Lower correlation was found in heterogeneous tumours if EB was used instead of DI (0.691 vs. 0.838). Good agreement with DI in AVE areas was observed in both homogenous and heterogeneous tumours (0.898 and 0.887). Concordance of tumour molecular profiles based on Ki-67 was better using DI in comparison to EB (Kappa index, 0.589 vs. 0.675). Whereas EB and DI were alike in homogeneous tumour, DI improved agreement in heterogeneous tumours, particularly in AVE areas. Subgroup analysis for tumour grades also showed improvement of correlation by DI in AVE areas in all G1/G2/G3 groups. Digital pathology using AVE method can be useful for Ki-67 scoring in daily practice, especially in heterogeneous and G2 tumours, by a substantial improvement of agreement between observers and results accuracy.

**Keywords** Digital analysis · Ki-67 · Breast cancer · Reproducibility · Immunohistochemistry · Prognosis

## Introduction

Ki-67 is a non-histone nuclear cortex protein, expressed in the cell nucleus during the G1, S, G2 and M phase

of the cell cycle, but not in the G0 (cell quiescent state) [1]. Ki-67 proliferative index (Ki-67) has become a predictive and prognostic factor in breast cancer (BC) [2]. According to St. Gallen's guidelines, high Ki-67 is one of the features that indicate increased risk of recurrence in oestrogen receptor (ER)-positive HER2-negative BC, indirectly supporting the value of adding chemotherapy to endocrine therapy in these patients [3–5].

St. Gallen 2013 consensus recommended using the Ki-67 index with a cut-off 20% in ER-positive HER2-negative BC whereas in 2015 a majority of the panel accepted a threshold value within the range of 20–29% [6, 7]. Certainly, the absence of a standardised methodology for Ki-67 immunostaining at the time makes Ki-67 assessment be both observer- and laboratory-dependent [4]. In spite of its value and introduction in clinical guidelines, the American Society of Clinical Oncology (ASCO) Tumour Marker Guidelines and the International Ki-67 Working Group do not recommend routine use of Ki-67 for prognosis in patients with newly diagnosed BC due to insufficient supporting clinical

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00428-018-2481-3>) contains supplementary material, which is available to authorized users.

---

✉ María del Rosario Taco Sanchez  
rtaco@bellvitgehospital.cat

<sup>1</sup> Department of Pathology, Bellvitge University Hospital, IDIBELL, CIBERONC, Barcelona, Spain

<sup>2</sup> Catalan Breast Unit, Catalan Institute of Oncology, ICO L'Hospitalet de Llobregat, Barcelona, Spain

<sup>3</sup> Department of Gynecology, Bellvitge University Hospital, IDIBELL, IRBLLEIDA, CIBERONC, Barcelona, Spain

<sup>4</sup> Bellvitge University Hospital and Arnau de Vilanova University Hospital, IDIBELL, IRBLLEIDA, CIBERONC, Barcelona, Spain

evidence, lack of standardisation and method reproducibility [8, 9]. Many studies have tried to define criteria to increase Ki-67 reproducibility [10]. The use of digital image analysis has been suggested as a potential method to improve the accuracy and inter-observer reproducibility in Ki-67 evaluation [11, 12].

Digital immunohistochemistry is one of the most promising applications brought by new generation image analysis. Whole-slide imaging systems convert glass slides into diagnostic quality digital images. While conventional immunohistochemistry interpretation is performed by semi-quantitative visual assessment using a microscope, automated digital immunohistochemistry is more precise and allows better definition of staining ranges that appear weak to the eye. Moreover, digital images allow the analysis of higher number of cells and produce continuous data, which may be subjected to multivariate data analysis.

The aim of this study is to evaluate the concordance between Ki-67 scoring assessed by two breast pathologists using two different methods, the eyeballing visual assessment on the microscope and the digital image analysis, assessed from hot spots and the average of hot and cold spots areas. The impact of these methods in clinical practice was also investigated.

## Material and methods

### Study population

A retrospective collection of Ki-67-immunostained (Ki-67-IHC) core needle biopsies (CNB) from BC patients diagnosed in 2010 at Bellvitge University Hospital (Spain) was selected for the study. All the slides were reviewed by two experienced breast pathologists (PatA and PatB) in a blinded manner. Unsatisfactory quality slides were discarded. The study was approved by the Ethical Committee Board of the hospital (code BEL-NUC-2014).

### Ki-67 index assessment

Ki-67-IHC was performed from sections of CNB fixed in formaldehyde for a period of time comprised between 6 and 72 h and embedded in paraffin. For the detection of the Ki-67 antigen, the mouse monoclonal antibody against pre-diluted DAKO human Ki-67 antigen (clone MIB-1, code IR626) was used. Antigen recovery was performed on “DAKO PT Link” with low pH solution (DAKO, K8005).

First, Ki-67 expression was categorised visually as homogeneous or heterogeneous based on the distribution of Ki-67-positive cells in the sections, by consensus between the two involved pathologists by visual interpretation. We defined the

tumour as homogeneous, when Ki-67 was similarly expressed throughout the tumour. Conversely, the definition of heterogeneous was established when areas of high and low proliferative indices were identified throughout the tumour. The percentage of Ki-67-positive cells was assessed for each case, considering nuclear staining of any intensity, including nucleolar staining.

For eyeballing visual assessment (EB) on the microscope, hot spot areas (HS) with high number of Ki-67-positive cells and cold spot areas with low proliferative index were identified using a low-power field ( $\times 20$ ) microscope. Ki-67 index was estimated by EB in both areas independently. The average between hot and cold spot areas was also calculated (AVE).

For digital image analysis (DI), Ki-67-IHC slides were scanned into digital images with the Panoramic MIDI scanner (3DHISTECH Ltd., Hungary) using the 20x objective. The NuclearQuant™ software was used for the automated evaluation of the Ki-67-IHC nuclear staining. Briefly, the software settings allow size and shape of tumour cells to be manually adjusted. Using colour deconvolution, the software detects cell nuclei and measures staining intensity on the chromogen channel, which can be adjusted to the stain protocol of the pathology lab. The software interpretation of digital images was visually inspected by the pathologists to ensure its accuracy. Then, the algorithm categorised the detected nuclei as negative and positive classes. As in EB, representative hot and cold spot areas were identified and selected within the digital images by both pathologists in a blinded manner avoiding non-neoplastic cells (lymphocytes, fibroblasts and endothelial cells) (Fig. 1). The same special settings of NuclearQuant™ software were used by both pathologists. The ratio of Ki-67 labelling index was independently obtained for hot and cold spots. The average of both regions was subsequently calculated.

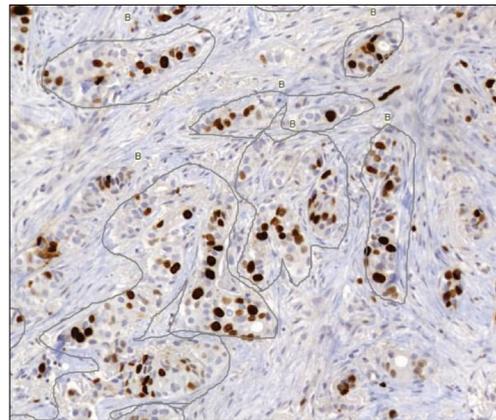
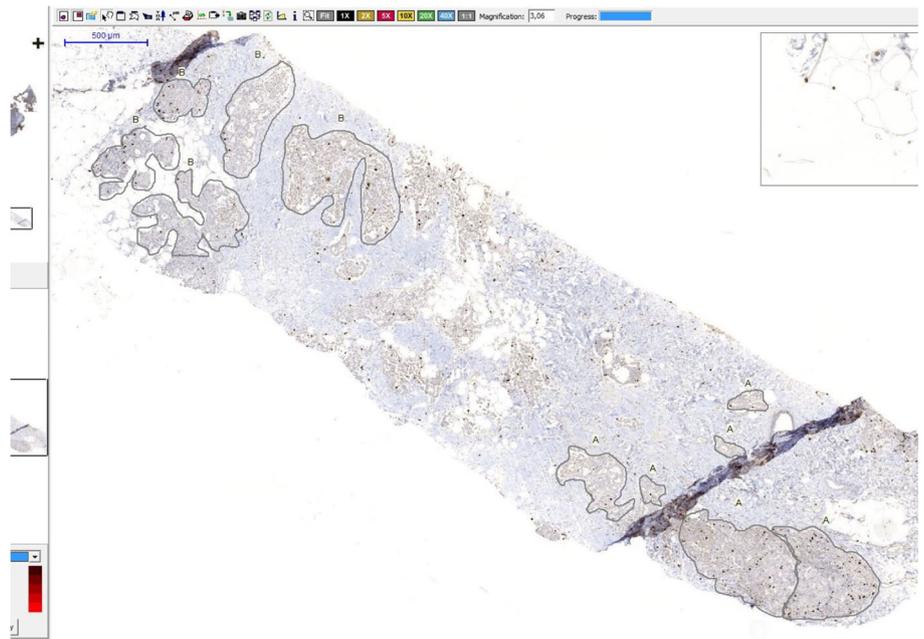
### Statistical analysis

Data analysis was performed using the SPSS Statistics package, Version 22.0 (Armonk, NY: IBM Corp and the MedCalc for Windows, version 15.0 (MedCalc Software, Ostend, Belgium). Categorical variables were described by absolute and relative frequencies, and chi-square test or Fisher's exact test was calculated. Continuous variables were analysed using centralisation, and dispersion statistics and Student's T or the Wilcoxon statistic was calculated. Two-tailed test and a significance level of 5% were applied.

Inter-observer and intra-observer intraclass correlation coefficient (ICC) of Ki-67 scores were calculated to assess the consistency between breast pathologists and EB and DI, stratified by HS and AVE areas. The 95% confidence interval (95%CI) was calculated in all cases.

Kappa statistics ( $K$ ) was calculated to evaluate inter-observer and intra-observer consistency of molecular profile reliant on Ki-67 scores according to St. Gallen 2013 recommendations [6] defined as luminal A for HER2-negative,

**Fig. 1** Ki-67 assessment in digital images. (A) Representative areas of hot and cold spots are selected, and labelled cells are quantified by NuclearQuat software. (B) An example of cold and hot spots selection



ER > 0, PR  $\geq$  20 and Ki-67 < 20% and luminal B for HER2-negative, ER > 0 and PR < 20 or Ki-67  $\geq$  20%. ICC was interpreted as follows [12, 13]: < 0.50 as “poor”, 0.50–0.74 as “moderate”. 0.75–0.90 as “good” and > 0.90 as “excellent”. Two-sided *p* values below 0.05 were considered statistical significant.

## Results

### Ki-67 index assessment

A total of 154 patients with CNB diagnosed of BC in 2010 were included in the study. Eighteen were rejected due to poor image quality. Regarding the 136 valid cases for the study, all patients were females, and the median age was 60 years, within the 26–86 range. Their pathological characteristics are depicted in Table 1. Tumours were mainly invasive carcinoma of no special type (NST), and the three tumour grades were similarly represented. Patients were mostly ER-positive (79.4%) and HER2-negative (85.5%), and most of them did not show lymphovascular invasion (92.6%). Most CNB had neoplastic cells in more than 50% of the tissue that was subjected to examination. Tumour homogeneity and heterogeneity were equally represented (50.7% and 49.3%).

Even though manual counting of as many as 500–1000 tumour cells has been recommended to evaluate Ki-67 index [9], we used EB at glance because it is a simpler and faster method although the total number of tumour cells is not assessed. In contrast, in DI, the total number of evaluated cells was recorded. The average of counted cell was balanced in all scenarios (PatA-HS, 1420 cells; PatB-HS, 1495 cells; PatA-AVE, 1399 cells; PatB-AVE, 1383cells).

#### 1. The mean and median Ki-67 indices are lower using DI in comparison with EB for both PatA and PatB in HS or AVE areas (Table 2)

In HS, EB showed a mean of Ki-67 index of 40.7% for PatA and 46.7% (22.5) for PatB. In DI, Ki-67 mean was 33.1% (21.8) and 38.1% (22.4) for PatA and PatB, respectively. Considering the AVE, Ki-67 had a mean of 33.4% (20.3) and 36.9% (21.4) in EB for PatA and PatB, and 28.9% (20.8) and 28.4% (19.4) in DI for PatA and PatB, respectively.

#### 2. Correlation between PatA and PatB was substantial using DI depending on assessment of HS or AVE areas (Table 3)

For EB, the correlation of Ki-67 scores between pathologists in HS was 0.828 (95%CI, 0.694–0.896) and 0.879 (0.815–0.919) in AVE areas. The Bland-Altman analysis showed a bias of the mean of  $-5.934$  (95% LOA,  $-29.8$ – $17.9$ ) in HS and  $-3.518$  (95%LOA,  $-23$ – $16$ ) in AVE.

**Table 1** Pathological characteristics of patients

Variables	N (%)	Mean (SD)	Median (range)
Tumour subtype			
No special type	113 (83.1)		
Lobular	19 (14.0)		
Others (mucinous, tubular or adenoid cystic)	4 (2.9)		
Histological grade (Nottingham grading system)			
G1	45 (33.1)		
G2	52 (38.2)		
G3	37 (27.2)		
Not Available	2 (1.5)		
Tumour tissue disposal		61.0 (20.5)	60.0 (10–90)
> 50%	89 (65.4)		
$\leq$ 50%	47 (34.6)		
Tumour heterogeneity			
Homogeneous	69 (50.7)		
Heterogeneous	67 (49.3)		
Lymphovascular invasion			
No	126 (92.6)		
Yes	7 (5.2)		
Not Available	3 (2.2)		
Oestrogen receptor		75.3 (39.6)	99.0 (0–100)
Positive (> 0%)	108 (79.4)		
Negative	28 (20.6)		
Progesterone receptor		49.4 (42.9)	55.0 (0–100)
Positive (> 0%)	94 (69.1)		
Negative	42 (30.9)		
HER2			
Positive	19 (14.0)		
Negative	116 (85.3)		
Not available	1 (0.7)		

SD, standard deviation

For DI, the correlation in HS was 0.861 (0.758–0.915) and 0.907 (0.873–0.933) in AVE. The Bland-Altman analysis showed a bias of the mean of  $-5.015$  (95%LOA,  $-26.35$ – $16.2$ ) in HS and 0.448 (95%LOA,  $-6.84$ – $17.74$ ) for AVE.

The correlation of Ki-67 scores between EB and DI of PatA was 0.878 (0.832–0.911) in HS and 0.912 (0.879–0.936) in AVE. Regarding PatB, the correlation between EB and DI was 0.839 (0.782–0.883) in HS and 0.825 (0.763–0.872) in AVE.

#### 3. Impact on molecular profiles (Tables S1–S2)

A total of 97 valid cases were used to assess the impact of the method used for Ki-67 assessment into the molecular profiles that depended on Ki-67 scores.

Regarding inter-observer impact, the agreement between pathologists in EB using HS examination was 90.7% and *K* was 0.589. When AVE was used, the agreement was 89.7%,

**Table 2** Ki-67 scores assessed by eyeballing and digital images stratified by observer and spot areas

		Eyeballing		Digital images	
		PatA	PatB	PatA	PatB
HS	Mean (SD)	40.7 (22.1)	46.7 (22.5)	33.1 (21.8)	38.1 (22.4)
	Median (range)	37.5 (5–90)	40.0 (10–95)	28.1 (2.3–86.7)	34.0 (3.2–94.8)
AVE	Mean (SD)	33.4 (20.3)	36.9 (21.4)	28.9 (20.81)	28.4 (19.4)
	Median (range)	28.7 (3–90)	31.2 (8–95)	23.5 (1.9–88.4)	23.5 (3.0–85.5)

HS, hot spots; AVE, average of hot and cold spots; PatA, pathologist A; PatB, pathologist B; SD, standard deviation

but  $K$  increased up to 0.725. In DI, the agreement between pathologists in HS was 86.5% with a  $K$  of 0.675. Using AVE, the agreement was 86.6%, and  $K$  was increased to 0.730. Having a moderate concordance between pathologists in HS using EB, coincidence was found over a half (8/15) of the cases catalogued as luminal A by PatA. On the contrary, concordance was improved using DI (0.675 vs. 0.589), increasing not only the number of Luminal A cases but also the coincidence [63% (21/33)]. Best agreement was reached using AVE with DI ( $K$ , 0.730), with higher agreement of luminal A categorisation in comparison with EB (38.1% vs. 19.6%).

Regarding intra-observer concordance, the agreement between EB and DI in HS of PatA was 81.4% with  $K$  of 0.524. Regarding AVE, the agreement was slightly reduced at 78.3%, but  $K$  increased to 0.557. For PatB, the agreement was 85.6% in HS assessment and 79.4% for AVE.  $K$  of molecular profiles in HS was 0.490 and increased up to 0.548 using AVE.

#### 4. Heterogeneity of the tumour (Table 4)

The correlation between pathologists in HS was very similar if tumour was classified as homogeneous in both EB and DI (0.869 and 0.866). In contrast, the agreement between pathologists decreased in tumours classified as heterogeneous (0.691 and 0.838, respectively). In both methods, the correlation improved in homogeneous tumours if AVE areas were considered instead of HS (EB, 0.887 vs. 0.869 and DI, 0.898 vs. 0.866). Interestingly, the use of AVE also increased correlation between pathologist for heterogeneous tumours in both EB (0.691 vs. 0.797) and DI (0.838 vs. 0.887) methods.

**Table 3** Inter-observer and intra-observer correlation of Ki-67 scores assessed by eyeballing and digital images stratified by spot areas. Data is shown as intraclass correlation coefficient (95% confidence interval)

	Pat A - PatB		EB - DI	
	EB	DI	Pat A	Pat B
HS	0.8282 (0.694–0.896)	0.8618 (0.758–0.915)	0.8780 (0.694–0.896)	0.8397 (0.782–0.883)
AVE	0.8791 (0.815–0.919)	0.9079 (0.873–0.933)	0.9124 (0.879–0.936)	0.8254 (0.763–0.872)

HS, hot spots; AVE, average of hot and cold spots; PatA, pathologist A; PatB, pathologist B; EB, eyeballing visual assessment; DI, digital images analysis

#### 5. Impact on histological grade (Table 5)

The correlation of Ki-67 scores was also inspected in HER2-negative cases according to histological grade subgroups and heterogeneity. DI improved concordance between pathologists in all tumour grade subgroups in global and also in both homogeneous and heterogeneous tumours (Table 5). Particularly in G2 grade, the worst reproducible group, DI reached good and better correlation in comparison with EB, with correlation over 0.800 if AVE areas were considered. Notably, the improvement in G2 tumours by DI with AVE areas in comparison with EB was substantial in both homogeneous and heterogeneous tumours (0.803 vs. 0.679 and 0.856 vs. 0.688, respectively).

#### Discussion

Ki-67 is a widely accessible biomarker of proliferation with strong evidence for clinical validity, prognosis and prediction in BC<sup>4</sup>. However, clinical application of Ki-67 is still under debate due to poor inter-observer reproducibility and lack of standardised IHC detection and scoring system [10], both stated in the guidelines of ASCO and International Ki-67 Working Group [8, 9].

Ki-67 index is usually defined as the percentage of positive tumour cell nuclei, counted in 3–10 HPF by testing 500–1000 tumour cells [1, 9]. Nonetheless, this method is monotonous, time-consuming and exhausting with a chance of leading to controversial results. In our study, we applied the EB since two recent studies shown EB estimation could be as good as the meticulous counting method being less time-consuming

**Table 4** Inter-observer correlation of Ki-67 scores assessed by eyeballing and digital images stratified by tumour heterogeneity. Data is shown as intraclass correlation coefficient (95% confidence interval)

		Homogeneous tumours	Heterogeneous tumours
HS	EB	0.869 (0.769–0.924)	0.691 (0.360–0.840)
	DI	0.866 (0.749–0.895)	0.838 (0.622–0.920)
AVE	EB	0.887 (0.779–0.939)	0.797 (0.685–0.872)
	DI	0.898 (0.836–0.937)	0.887 (0.821–0.930)

HS, hot spots; AVE, average of hot and cold spots; EB, eyeballing visual assessment; DI, digital images analysis

[14, 15]. Therefore, the total number of cells counted by microscope examination was not available. In contrast, total cell counting was available using DI, not only fulfilling the criteria for total cell count but also beyond the range of cell amounts. Our results suggest that DI is a potential efficient method for Ki-67 assessment with better performance in terms of capacity, precision and accuracy in comparison with EB or cell counting [16–19].

We found lower Ki-67 values using DI in comparison with EB of tumour area examined (HS or AVE) for both pathologists independently. These results are in line with those obtained by Maeda's group, who reported a mean of 22% ( $\pm$  12.8%) Ki-67 index by EB and 20.4% ( $\pm$  12.7) by DI. Our higher mean values could be given as we consider Ki-67-positive cells whether any intensity of nuclear staining was observed.

In our study, we aimed to investigate the correlation of two methods, EB and DI, for Ki-67 index assessment in BC, and also to evaluate the effect of specific tumour areas examined, HS and AVE. Our results showed that correlation between pathologists in HS was slightly better using DI instead of EB. Indeed, the agreement in DI increased if AVE areas were considered instead of HS. These results are in line with those reported by Zhong et al. who also found a good agreement between pathologists using DI and AVE [11].

We also found similar concordance between pathologists if tumour was classified as homogeneous, independent of Ki-67 assessment method (EB or DI). Nevertheless, the agreement was reduced in tumours classified as heterogeneous, especially using EB, as other studies have shown [1, 9, 11]. Therefore, our results suggest that AVE may minimise the effect of tumour heterogeneity on Ki-67 scoring, with significant higher correlation between observers using DI. In our opinion, DI plus AVE methodology could contribute to standardising the analysis of Ki-67 index and optimising a cut-off in each pathology laboratory.

The lack of standardised methods to evaluate Ki-67 index entails inter-observer variability that may affect therapy decision making as for many other protein markers whose expression is quantified in tumour sections. Ki-67 assessment is important in distinguishing luminal A from luminal B molecular profiles. Inappropriate molecular profile designation may have consequences in BC patients' systemic treatment personalization as St. Gallen's recommendations states [6]. In our study, the concordance of molecular profiles between pathologists was moderate if HS were examined. However, concordance increased using AVE areas, and more importantly, DI provided the highest concordance between pathologists. In our opinion, Ki-67 assessment by can DI minimises the impact of inter-observer variation in molecular profiling designation and also the risk of inappropriate treatment decision. Nevertheless, further studies are needed to investigate more accurately the impact on patient management according to Ki-67 assessment with DI. Prospective studies must be addressed to evaluate follow-up and clinical response outcomes of patients whose treatment was tailored according to digital scoring of Ki-67.

On the other hand, even though we did not focus our investigations in time savings, we believe that DI may significantly increase the efficiency of Ki-67 interpretation, reducing the turnaround time for Ki-67 scoring.

Currently, the International Group recommends the visual quantification of Ki-67 by microscopy as a gold standard

**Table 5** Inter-observer correlation of Ki-67 scores assessed by eyeballing and digital images in HER2-negative patients stratified by tumour grades (G1–G3). Data is shown as intraclass correlation coefficient (95% confidence interval)

		General			Homogenous tumours			Heterogeneous tumours		
		G1 (n = 45)	G2 (n = 52)	G3 (n = 37)	G1 (n = 32)	G2 (n = 29)	G3 (n = 8)	G1 (n = 13)	G2 (n = 23)	G3 (n = 29)
HS	EB	0.718 (0.557–0.839)	0.680 (0.523–0.812)	0.812 (0.660–0.917)	0.536 (0.252–0.742)	0.684 (0.409–0.874)	0.895*	0.882*	0.639 (0.353–0.835)	0.729 (0.369–0.889)
	DI	0.845 (0.722–0.921)	0.727 (0.528–0.868)	0.861 (0.709–0.940)	0.844 (0.650–0.939)	0.726 (0.529–0.869)	0.931*	0.854*	0.714 (0.351–0.941)	0.815 (0.583–0.969)
AVE	EB	0.739 (0.589–0.835)	0.717 (0.531–0.847)	0.869 (0.755–0.943)	0.580 (0.294–0.764)	0.679 (0.486–0.845)	0.940*	0.885*	0.688 (0.505–0.866)	0.816 (0.673–0.919)
	DI	0.824 (0.721–0.934)	0.843 (0.703–0.927)	0.874 (0.739–0.953)	0.851 (0.673–0.945)	0.803 (0.620–0.931)	0.979*	0.840*	0.856 (0.645–0.924)	0.845 (0.688–0.946)

HS, hot spots; AVE, average of hot and cold spots; EB, eyeballing visual assessment; DI, digital images analysis. NA, not available

\*No significant values due to small sample size of subgroups

technique. This can be simple in homogeneous tumours growing in small nests but challenging in heterogeneous tumours as optical quantification can be really difficult in tumours that grow with solid big nests. In our opinion, Ki-67 assessment by EB can be appropriate in homogeneous tumours. However, we think DI may be particularly helpful in heterogeneous tumours since, in our results, good agreement between pathologists was reached using DI in both heterogeneous and homogeneous tumours. Indeed, DI allows pathologists to quantified higher amount of cells in least time and different areas, and covers a higher tumour section. Moreover, DI easily allows the normalisation of tumour areas subjected to study (AVE method), which is very important for continuous variables such as Ki-67 scoring, having a significant impact on the accuracy of the results, particularly in heterogeneous tumours.

Moreover, the results of the analysis of tumour grade subgroups corroborate better performance of Ki-67 scoring agreement between pathologists using DI with AVE methodology, notably the improvement of pathologists' agreement in G2 tumours, the worst reproducible subgroup [20]. Even though further studies should be addressed to confirm these results, our results point out the potential benefits of DI for standardised Ki-67 index assessment in pathology laboratories.

In conclusion, digital immunohistochemistry is a good tool to improve reproducibility of Ki-67 scoring. In our study, we found a better correlation between pathologists using DI than EB even for molecular profiles Ki-67-dependent. The DI turns into a more accurate assessment, particularly for heterogeneous and G2 tumours using AVE scores. Nevertheless, further prospective studies are needed to evaluate the impact of Ki-67 digital assessment on BC patients' treatment management.

**Acknowledgements** We thank Xavier Pérez Martín (Clinical Research Unit, Catalan Oncology Institute) for the statistical analysis support. We also thank Sysmex España for technical support.

**Ethical responsibilities of authors section** All authors contributed equally to the study design, results analysis and manuscript conception.

## Compliance with ethical standards

The study was approved by the Ethics Committee of the hospital. The study was performed in compliance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA (2010) Ki-67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* 11:174–183. [https://doi.org/10.1016/S1470-2045\(09\)70262-1](https://doi.org/10.1016/S1470-2045(09)70262-1)
- Penault-Llorca F, Radošević-Robin N (2016) Biomarkers of residual disease after neoadjuvant therapy for breast cancer. *Nat Rev Clin Oncol* advance on:487–503. <https://doi.org/10.1038/nrclinonc.2016.1>
- Esposito A, Criscitiello C, Curigliano G (2015) Highlights from the 14(th) St Gallen International Breast Cancer Conference 2015 in Vienna: dealing with classification, prognostication, and prediction refinement to personalize the treatment of patients with early breast cancer. *Ecancermedicalscience* 9:518. <https://doi.org/10.3332/ecancer.2015.518>
- Focke CM, van Diest PJ, Decker T (2016) St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki-67-based proliferation assessment methods. *Breast Cancer Res Treat* 159:257–263. <https://doi.org/10.1007/s10549-016-3950-5>
- Criscitiello C, Disalvatore D, De Laurentis M et al (2014) High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in luminal B HER2 negative and node-positive breast cancer. *Breast* 23:69–75. <https://doi.org/10.1016/j.breast.2013.11.007>
- Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ, Albain KS, André F, Bergh J, Bonnefoi H, Bretel-Morales D, Burstein H, Cardoso F, Castiglione-Gertsch M, Coates AS, Colleoni M, Costa A, Curigliano G, Davidson NE, di Leo A, Ejlertsen B, Forbes JF, Gelber RD, Gnant M, Goldhirsch A, Goodwin P, Goss PE, Harris JR, Hayes DF, Hudis CA, Ingle JN, Jassem J, Jiang Z, Karlsson P, Loibl S, Morrow M, Namer M, Kent Osborne C, Partridge AH, Penault-Llorca F, Perou CM, Piccart-Gebhart MJ, Pritchard KI, Rutgers EJT, Sedlmayer F, Semiglazov V, Shao ZM, Smith I, Thürlimann B, Toi M, Tutt A, Untch M, Viale G, Watanabe T, Wilcken N, Winer EP, Wood WC (2013) 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. *Ann Oncol* 24:2206–2223. <https://doi.org/10.1093/annonc/mdt303>
- Gnant M, Thomssen C, Harbeck N (2015) St. Gallen/Vienna 2015: a brief summary of the consensus discussion. *Breast Care* 10:124–130
- Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC Jr, American Society of Clinical Oncology (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25:5287–5312. <https://doi.org/10.1200/JCO.2007.14.2364>
- Dowsett M, Nielsen TO, A'Hern R et al (2011) Assessment of Ki-67 in breast cancer: recommendations from the international Ki-67 in breast cancer working group. *J Natl Cancer Inst* 103:1656–1664. <https://doi.org/10.1093/jnci/djr393>
- Polley M-YC, Leung SCY, Gao D et al (2015) An international study to increase concordance in Ki-67 scoring. *Mod Pathol* 28:778–786
- Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R (2016) A comparison of visual assessment and automated digital image analysis of Ki-67 labeling index in breast cancer. *PLoS One* 11:e0150505. <https://doi.org/10.1371/journal.pone.0150505>
- Fasanella S, Leonardi E, Cantaloni C, Eccher C, Bazzanella I, Aldovini D, Bragantini E, Morelli L, Cuorvo LV, Ferro A, Gasperetti F, Berlanda G, Dalla Palma P, Barbareschi M (2011) Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies. *Diagn Pathol* 6(Suppl 1):S7. <https://doi.org/10.1186/1746-1596-6-S1-S7>
- Koo TK, Li MY (2016) A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, Obermann E, Ohlschlegel C, Padberg B, Rakozy C, Sancho Oliver S, Schobinger-Clement S, Schreiber-Facklam H, Singer G, Tapia C, Wagner U, Mastropasqua MG, Viale G, Lehr HA (2012)

- How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss working group of breast- and gynecopathologists. *PLoS One* 7:1–12. <https://doi.org/10.1371/journal.pone.0037379>
15. Vörös A, Csörgő E, Kővári B, Lázár P, Kelemen G, Ruzs O, Nyári T, Csemi G (2015) Different methods of pretreatment Ki-67 labeling index evaluation in Core biopsies of breast Cancer patients treated with neoadjuvant chemotherapy and their relation to response to therapy. *Pathol Oncol Res* 21:147–155. <https://doi.org/10.1007/s12253-014-9800-z>
  16. Yoshiki M, Takayuki U, Kenichi Y et al (2013) Interobserver concordance of Ki-67 labeling index in breast cancer: Japan Breast Cancer Research Group Ki-67 ring study. *Cancer Sci* 104:1539–1543. <https://doi.org/10.1111/cas.12245>
  17. Einar G, Ivar S, M JEA et al (2012) Comparison of the effect of different techniques for measurement of Ki-67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. *Histopathology* 61:1134–1144. <https://doi.org/10.1111/j.1365-2559.2012.04329.x>
  18. Konsti J, Lundin M, Joensuu H, Lehtimäki T, Sihto H, Holli K, Turpeenniemi-Hujanen T, Kataja V, Sailas L, Isola J, Lundin J (2011) Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer. *BMC Clin Pathol* 11:3. <https://doi.org/10.1186/1472-6890-11-3>
  19. Mohammed ZMA, McMillan DC, Elsberger B et al (2012) 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* 106:383–388. <https://doi.org/10.1038/bjc.2011.569>
  20. Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, Obermann E, Ohlschlegel C, Padberg B, Rakozy C, Sancho Oliver S, Schobinger-Clement S, Schreiber-Facklam H, Singer G, Tapia C, Wagner U, Mastropasqua MG, Viale G, Lehr HA (2012) How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss working group of breast- and gynecopathologists. *PLoS One* 7:e37379. <https://doi.org/10.1371/journal.pone.0037379>