The Breast 49 (2020) 225–232

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

# An interobserver reproducibility analysis of size-set semiautomatic counting for Ki67 assessment in breast cancer

Yi-xing Wang [a], Yuan-yuan Wang [a], Cheng-gang Yang [b], Hong Bu [c], Wen-tao Yang [d], Li Wang [a], Wen-mang Xu [a], Xi-long Zhao [a], Wen-xing Zhao [a], Lei Li [b], Shu-ling Song [a], Ju-lun Yang [a, *]

[a] Department of Pathology, 920th Hospital of Joint Logistics Support Force of PLA, Kunming, Yunnan, 650100, PR China
[b] Department of Pathology, Yunnan Cancer Hospital, Kunming, Yunnan, 650032, PR China
[c] Department of Pathology, West China Hospital, Sichuan University, Chengdu, 610041, PR China
[d] Department of Pathology, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, 200032, PR China

ARTICLE INFO

*Article history:*
Received 20 August 2019
Received in revised form
12 December 2019
Accepted 13 December 2019
Available online 20 December 2019

*Keywords:*
Ki67 label index
Breast cancer
Immunochemistry
Assessment
Reproducibility

ABSTRACT

*Purpose:* The proliferation marker Ki67 has prognostic and predictive values in breast cancer, and the cutoff of the Ki67 label index (LI) is a key index for chemotherapy. However, poor interobserver consistency in Ki67 assessment has limited the clinical use of Ki67, especially in luminal cancers. Here, we reported a modified Ki67 assessment method, size-set semiautomatic counting (SSSAC) and investigated its interobserver reproducibility.

*Methods:* One hundred invasive breast cancer tissues were set immunostained for Ki67 in one laboratory, scanned as digital slides, and sent to 41 pathologists at the laboratories of 16 hospitals for Ki67 LI assessment using size-set semiautomatic counting (SSSAC), size-set visual assessment (SSVA) and size-set digital image analysis (SSDIA) with a specific image viewing software (Aperio Image Scope, Leica, Germany). The intraclass correlation coefficient (ICC) and Bland-Altman plot were used to evaluate interobserver reproducibility. The Wilcoxon signed-rank test was used to analyze the difference in the Ki67 values assessed by SSSAC and SSDIA.

*Results:* SSSAC demonstrated better interobserver reproducibility (ICC = 0.942) than SSVA (ICC = 0.802). The interobserver reproducibility was better in Ki67 homogeneously stained slides and centralized hot-spot slides than in scattered hot-spot slides. The Ki67 value assessed with SSSAC was obviously higher than that assessed with SSDIA (negative ranks (SSDIA < SSSAC): N = 80, sum of ranks = 4274.50; positive ranks (SSDIA > SSSAC): N = 17, sum of ranks = 478.50; Z = −6.837; P < 0.001).

*Conclusion:* SSSAC shows satisfactory interobserver reproducibility in the Ki67 assessment of breast cancer and may be a candidate standard method for Ki67 LI assessment in breast cancer and other malignancies.



## 1. Introduction

Ki67 is universally expressed in proliferating cells and has become a nuclear proliferation marker of malignant tumors [1,2]. Several studies have demonstrated the prognostic value of the Ki67 label index (LI) for breast cancer patients [3–6]. Potential uses of Ki67 LI include the prediction of the relative responsiveness or resistance to chemotherapy, estimation of the residual risk in patients on standard therapy and as a dynamic biomarker of treatment efficacy in samples taken before, during, and after neoadjuvant therapy, particularly neoadjuvant endocrine therapy [7]. Currently, the assessment of Ki67 using immunohistochemical

* Corresponding author. Department of Pathology, 920th Hospital of Joint Logistics Support Force of PLA, 212 Daguan Road, Xishan District, Kunming, Yunnan, China.

*E-mail addresses:* 784503159@qq.com (Y.-x. Wang), wyy131328@126.com (Y.-y. Wang), 2399853539@qq.com (C.-g. Yang), hbpathology@126.com (H. Bu), yangwt2000@163.com (W.-t. Yang), 2001wl@163.com (L. Wang), highskysky@126.com (W.-m. Xu), zhaoxilong@aliyun.com (X.-l. Zhao), wenxingzhao@sohu.com (W.-x. Zhao), 120.lilei@163.com (L. Li), 871819779@qq.com (S.-l. Song), yangjulun@sina.com (J.-l. Yang).

https://doi.org/10.1016/j.breast.2019.12.009

(IHC) staining in breast cancer subtypes has become widespread.

However, the assessment of Ki67 has poor reproducibility [8]. More than one factor leads to the poor reproducibility, such as scoring area, scoring method, and variation in the evaluated fields under microscopes [9]. Although a formal standard method for Ki67 LI assessment have been proposed by the International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group [7], to date there is low reproducibility. In clinical practice, pathologists commonly use visual assessment ('quick-scan rapid Ki67 estimate') or manual counting under the microscope to evaluate Ki67 LI. However, visual assessment lacks interobserver reproducibility [9–11]. In manual counting, at least 500–1000 tumor cells have to be counted to achieve acceptable error rates and correct for heterogeneity, a time-consuming and error-prone process [7,12]. Digital image analysis (DIA) may be a better candidate for Ki67 assessment, but software does not possess the ability to recognize the region of interest accurately and distinguish every type of breast cancer cell exactly [13,14].

Therefore, a well operable and good reproducible Ki67 assessment method is crucial and urgently needed for the precise therapy of breast cancer. In this study, we report a modified Ki67 assessment method, size-set semiautomatic counting (SSSAC) and investigated the interobserver reproducibility among 41 pathologists at 16 hospitals.

## 2. Materials and methods

### 2.1. Case selection

One hundred cases of primary invasive breast cancer paraffin-embedded tissue blocks from surgical specimens were randomly selected from the pathology databases of 920th Hospital of Joint Logistics Support Force of PLA and The Third Affiliated Hospital of Kunming Medical University during the years of 2014 and 2015. All of them are not otherwise specified (NOS) invasive ductal carcinoma,. All the patients received no neoadjuvant chemotherapy before undergoing surgery. All the samples in this study were further consulted by two pathologists according to the standard of World Health Organization [15].

### 2.2. Immunohistochemical staining and image acquisition

Tissue specimens were fixed with 4% neutral buffered formalin and were embedded in paraffin. The paraffin-embedded tumor tissues were cut into 3-μm-thick sections. To minimize the influence caused by the thickness of the tissue sections, all the tumor tissues blocks were cut into 3-μm sections on the same microtome with a standardized speed by one technician. Tissue sections were immunostained with antibody MM1 (Novocastra, Newcastle, U.K.) using the Leica Bond-Max staining robot (Leica Microsystems, Bannockburn, IL, USA) at one laboratory. The 100 qualified slides were digitally scanned into digital slides at × 20 magnification using the Aperiod AT2 digital scanner (Leica Biosystem, Wetzlar, Germany) and then were sent to 41 pathologists at 16 hospitals for Ki67 LI assessment.

### 2.3. Size-set semiautomatic counting (SSSAC)

First, hot spots of Ki67 staining were determined. The nonconformity of scoring areas is a main factor that affects the reproducibility of Ki67 assessment. International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group recommends that, if there are clear hot spots, data from these should be included in the overall score [7]. Gudlaugsson E et al. showed that Ki67 in the hot spots is of utmost prognostic significance compared with the average and lowest zones [11]. Accordingly, we defined a hot spot where Ki67 staining is particularly prevalent as the scoring area. In heterogeneous slide, the fields of Ki67-positive tumor cells include high, medium, low staining areas and negligible areas. The high staining area was defined as hot spot [16–18]. The digital slides were observed at × 10 magnification to identify hot-spot areas. If the hot spots were distributed in the invasive edge of the tumor, the selection of hot spots should include the invasive edge of the tumor. If the staining of the slides was homogeneous, the scoring area could be any area on the whole slide.

Next, the scoring frames were demarcated. On the area of the hot spot, three 200 × 200-μm square frames were demarcated (Fig. 1A). Demarcation of the scoring frames in the hot spots were performed as follows: (1) demarcation of three frames successively if the area of the hot spot was sufficiently large; (2) in some slides, the hot spots were insufficient to set three frames—in this circumstance, three frames were demarcated in different hot spots; (3) if the staining was homogeneous, three frames were demarcated randomly; (4) after demarcating three frames in the hot spots, the frames were moved slightly to ensure the Ki67 staining in areas limited by the frames were the highest. Additionally, three frames could not overlap with each other. If the number of cancer cells in the three scoring frames is less than 500, another scoring frame will be demarcated in the hot spot area until the number of cancer cells exceeds 500 (Fig. 1B).

Finally, the percent of Ki67-positive tumor cells (Ki67 LI) on the Ki67-immunostained digital slides was semiautomatically counted using the digital image viewing software (Aperio Image Scope, Leica, Germany). Only nuclear brown staining is considered Ki67-positive. Staining intensity is not relevant. Tumor cells with blue counterstained nucleus was defined as Ki67-negative cells. Briefly, Ki67-positive tumor cells and negative cells covered by three scoring frames were counted at appreciate magnification by clicking the tumor cells using the mouse button manually. When the mouse button was clicked on target cells, the software covered cells with a crucial marker to avoid counting cells twice, and ordinal numbers were presented on the marker automatically at the same time (Fig. 1C). When all the Ki67-positive tumor cells were clicked, the total number of Ki67-positive tumor cells was presented in the annotation layer. The Ki67-negative tumor cells were counted by the same method. Finally, the Ki67 LI was defined as [Ki67-positive cells/(Ki67-positive cells + Ki67-negative cells)] × 100.

### 2.4. Size-set visual assessment (SSVA)

The Ki67 LI of three demarcated scoring frames mentioned above was gained by glancing rapidly by pathologists at × 40 magnification at a digital slide. Brown nuclei rather than blue were scored as positive. The values of Ki67 LI were evaluated at 1% intervals in cases in which the Ki67 LI values were <10% or >90%, and the values of Ki67 LI were evaluated at 5% intervals in cases in which the Ki67 LI values were 10%–90%.

### 2.5. Size-set digital image analysis (SSDIA)

The percent values of Ki67-positive cells in the three frames were evaluated automatically using digital image analysis (DIA) software (Aperio Nuclear Algorithm; Leica Biosystems, Germany) referring to manufacturer's manual. Briefly, tumor cell parameters based on the size, smoothing, merging, trimming, roundness, compactness and elongation of the nuclei were developed to distinguish tumor cells from interstitial cells (fibroblasts, fibrocyte, lymphocyte cells, histocytes, endoepitheliocytes). We debugged

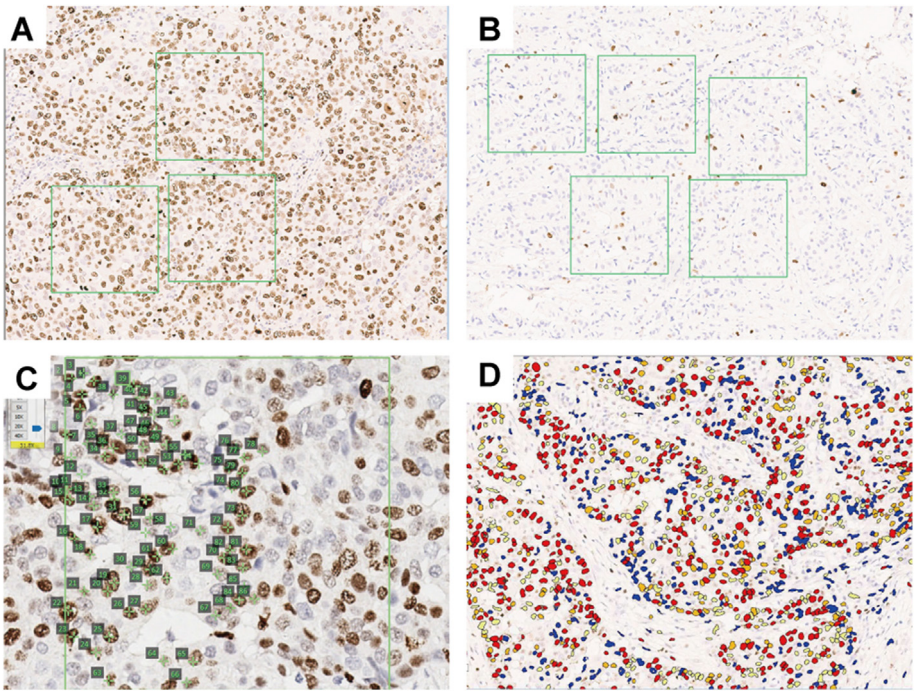


**Fig. 1.** Ki67 assessment of breast cancer. A. Demarcated three scoring frames with the size of 200 × 200 μm on the hot spot. B. More than three scoring frames were demarcated when the cancer cells in three scoring frames were less than 500. C. SSSAC assessment. When clicking on a cancer cell manually the image viewing software covers the cell with a crucial marker and ordinal numbers were presented on the marker automatically. D. SSDIA assessment. The size of 25μm$^2$ was set as cut off of nucleus to separate the interstitial cells from breast cancer cells. Ki67 immunostaining of breast cancer cells was set as strong (Red), moderate (Orange), weak (yellow) and negative (Blue).

different tumor cell parameters repeatedly according to the software instructions until almost all of the interstitial cells were separated from cancer cells. (Fig. 1D). Then, the scoring criteria was modified to define Ki67-negative and Ki67-positive tumor cell nuclei segmented, Ki67 immunostaining of breast cancer cells was set as strong (Red), moderate (Orange), weak (yellow) and negative (Blue), and tumor cells with brown cytoplasm were excluded. Finally, three square frames with a size of 200 × 200 μm were demarcated on the hot spots of each digital slide, and the Ki67 LI of cancer cells demarcated in the three scoring frames were calculated automatically by the Aperio Nuclear Algorithm image analysis software.

### 2.6. Pathologists

Total 41 pathologists participated in the assessment of Ki67. Among them, 35 were experienced pathologists who engaged in pathological diagnosis for more than 5 years and 6 inexperienced ones engaged in pathological diagnosis for less than 5 years. 32 came from provincial hospital laboratory and 9 worked in state hospital laboratory.

### 2.7. Statistical analysis

To measure the interobserver reproducibility of the Ki67 LI assessment, the intraclass correlation coefficient (ICC) was estimated with a 95% confidence interval (CI) using two-way mixed models. The ICC has a range of 0–1, with the consensus that the closer distance is to 1, the higher the agreement will be. The ICC demonstrates no universally accepted standard criteria. Hence, based on the similarity to the kappa coefficient, the following criteria were used here to aid interpretation [19,20]: <0.4 indicates "bad correlation"; 0.4–0.69 indicates "moderate correlation"; 0.7–0.79 indicates "substantial correlation"; >0.80 indicates "almost perfect correlation".

The Bland-Altman plot was used to reveal the difference in the Ki67 LI intuitively between two paired pathologists using the methods of SSVA and SSSAC. If there is high agreement, the differences are expected to be centered about the middle solid line, with a small standard deviation [21].

The difference between SSSAC and SSDIA values was analyzed using the Wilcoxon signed-rank test.

The ICC and Wilcoxon signed-rank analyses were performed using the software package SPSS 22.0, and Bland-Altman Plot was generated using the software package MedCalc 15.2.2.

## 3. Results

### 3.1. Pattern of Ki67 immunostaining

In 100 cases of breast cancer, 83 cases displayed heterogenous Ki67 immunostaining, that is, there were Ki67 hot spots in these cases (Fig. 2A), and the remaining 17 cases showed homogeneous distributions of Ki67-positive tumor cells (There were no Ki67 hot spots) (Fig. 2B). The shapes of the hot spots were varied, such as being round, oval, and irregular. The size, location and distribution of the hot spots are shown in Table 1. The maximum diameter of the hot spot was 11,740 μm, and the minimum diameter was 50 μm. The diameters of 64% cases were more than 200 μm. The distribution and location of the hot spots were distinct in different cases. The hot spots of 32 cases were distributed centrally (Fig. 2C), and 51 cases were distributed in a scattered manner (Fig. 2D). The hot spots were located in the invasive edge of the tumor in 29 cases (Fig. 2E) and in the interior of the tumor in 40 cases (Fig. 2F); additionally, the hot spots of the other 14 cases were located both in the invasive edge and interior of the tumor.

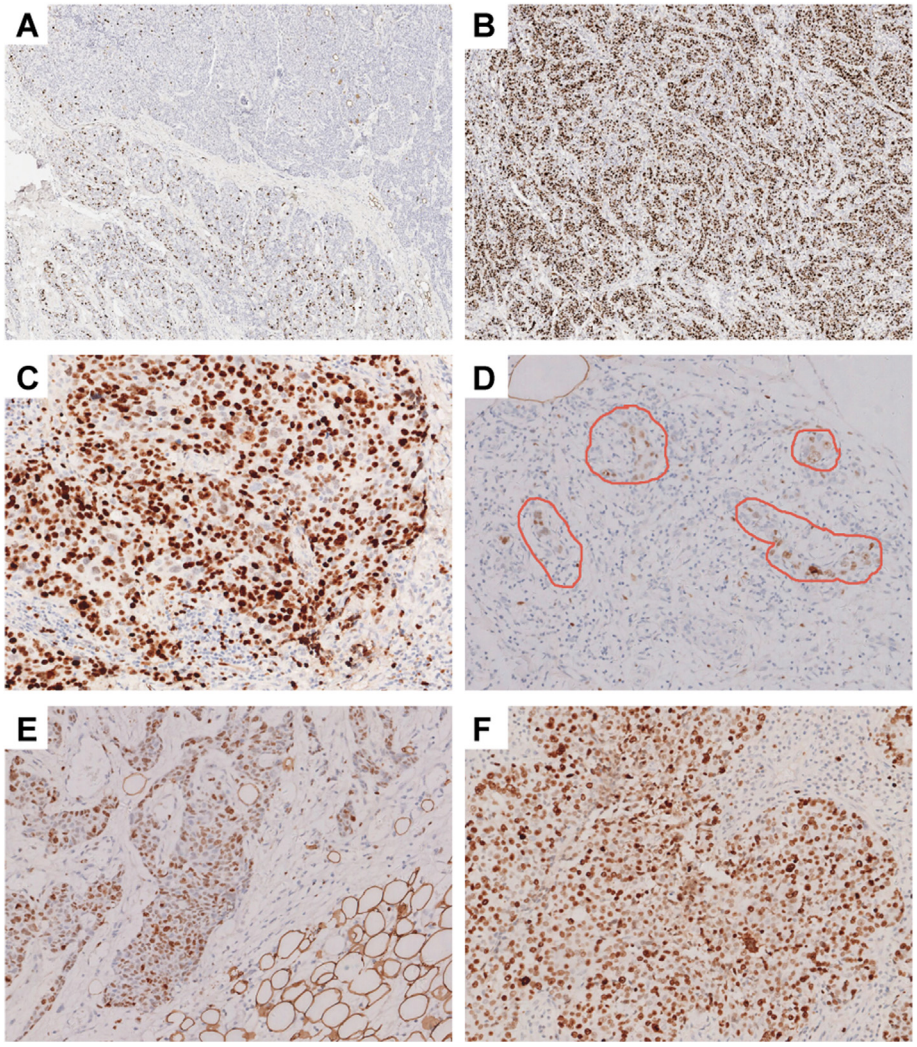


**Fig. 2.** The pattern of Ki67 immunostaining in breast cancer. A. Heterogenous staining of Ki67. There was a hot spot in the lower-left corner. B. Homogenous staining. C. The hot spot distributed centrally. D. The areas annotated by red circle were scattered hot spots. E. The hot spot in the invasive edge of the tumor. F. The hot spot in the interior of the tumor. A high-resolution version of the image is available as eSlide:VM00001-VM00006.

**Table 1**
Size, location and distribution of the hot spots.

| Size of hot spot (diameter) | n | Percentage (%) | Distribution of hot spot (n) | | Location of hot spot (n) | | |
|---|---|---|---|---|---|---|---|
| | | | centralized | scattered | invasive edge of the tumor | interior of the tumor | interior and edge of the tumor |
| <100 μm | 13 | 16 | 0 | 13 | 1 | 9 | 3 |
| 100–200 μm | 17 | 20 | 1 | 16 | 4 | 9 | 4 |
| >200 μm | 53 | 64 | 31 | 22 | 24 | 22 | 7 |

### 3.2. Size of the scoring frames

We counted the tumor cells covered by different-sized frames and found that each of the 100 × 100-μm, 200 × 200-μm or 300 × 300-μm scoring frames could cover approximately 50, 200 or 600 tumor cells, respectively. Thus, three 200 × 200-μm scoring frames could cover approximately 600 tumor cells (at least 500 tumor cells were required by International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group [9]). If 100 × 100-μm scoring frames were used, 10 frames must be demarcated to satisfy the requirement of International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group, complicating the Ki67 LI assessment. The 300 × 300-μm scoring frames had a greater possibility to cover non-hot spots around hot spots compared with the 200 × 200-μm scoring frames. Thus, 200 × 200-μm scoring frames were selected in our study.

### 3.3. Comparison of SSSAC and SSVA/SSDIA

ICC analysis revealed that SSSAC had higher agreement in Ki67 assessment of breast cancer among inter-pathologists than SSVA. The Ki67 LI obtained by SSSAC had almost perfect concordance among 41 pathologists (ICC, 0.942 [95% CI: 0.926, 0.957]) (Table 2).

**Table 2**
ICC of Ki67 LI in different groups.

| Groups | n | ICC (95% CI) | |
|---|---|---|---|
| | | SSVA | SSSAC |
| Pathologists | | | |
| All | 41 | 0.942 (0.926, 0.957) | 0.802 (0.756, 0.846) |
| Experienced pathologists | 35 | 0.805 (0.760, 0.849) | 0.953 (0.940, 0.965) |
| Inexperienced pathologists | 6 | 0.786 (0.730, 0.837) | 0.903 (0.873, 0.928) |
| Laboratory | | | |
| Provincial laboratory | 32 | 0.801 (0.755, 0.845) | 0.946 (0.930, 0.960) |
| The state laboratory | 9 | 0.812 (0.764, 0.856) | 0.929 (0.907, 0.947) |
| Distribution of Ki67-positive tumor cells | | | |
| homogeneous | 17 | 0.876 (0.791, 0.945) | 0.967 (0.940, 0.986) |
| centralized hot spot | 32 | 0.745 (0.670, 0.820) | 0.949 (0.922, 0.970) |
| scattered hot spot | 51 | 0.648 (0.560, 0.742) | 0.862 (0.812, 0.906) |
| Grade | | | |
| 1 | 8 | 0.567 (0.352, 0.847) | 0.955 (0.901, 0.989) |
| 2 | 86 | 0.943 (0.862, 0.990) | 0.947 (0.931, 0.962) |
| 3 | 6 | 0.761 (0.704, 0.816) | 0.969 (0.922, 0.995) |

To further compare the reproducibility of interobserver using SSSAC and SSVA, we randomly selected the Ki67 LI assessment data of 12 pathologists from 41 pathologists, divided into 6 paired groups and compared their Ki67 LI evaluation results with both SSVA and SSSAC using Bland-Altman plots. Six paired Bland-Altman plots intuitively showed that differences were more concentrated about the middle solid line when using SSSAC than using SSVA, demonstrating that SSSAC has higher reproducibility than SSVA (Fig. 3).

When the parameters of digital image analysis are set, almost the same results will be produced for the same image. Therefore,

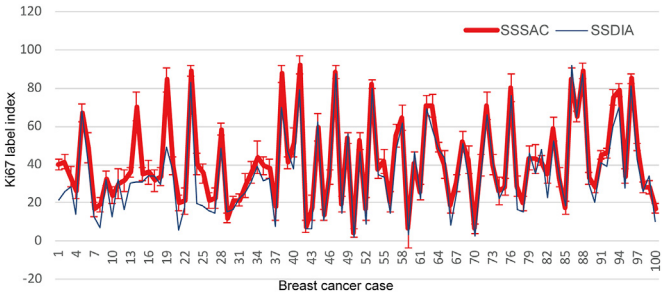


**Fig. 4.** Comparison of Ki67 LI in 100 cases of breast cancer between SSSAC and SSDIA assessment (. There were mean Ki67 LI and SD for each case from pathologists in SSSAC line. The Ki67 LI evaluated by SSSAC was larger than that by SSDIA in majority of cases.

DIA could provide precise reproducible analyses. However, it has its own bias and produces inaccurate data. To compare the difference in the Ki67 LI assessed with SSSAC and SSDIA, the mean Ki67 LI of 35 experienced pathologists from 41 pathologists using SSSAC was calculated as the final SSSAC value and then was compared with that of SSDIA data. In 80 cases of breast cancer the Ki67 LI obtained by SSSAC was higher than that obtained by SSDIA. In 17 cases, the Ki67 LI obtained by SSSAC was smaller than that by SSDIA. In three cases, the Ki67 LI obtained by SSSAC was equal to that by SSDIA (Fig. 4). The maximum difference in the Ki67 LI between SSSAC and SSDIA was 18, and the minimum difference between them was 0. To further demonstrate the difference in the Ki67 value assessed by SSSAC and SSDIA, the data were analyzed by the Wilcoxon signed-rank test, which showed that the Ki67 LI with SSSAC was obviously

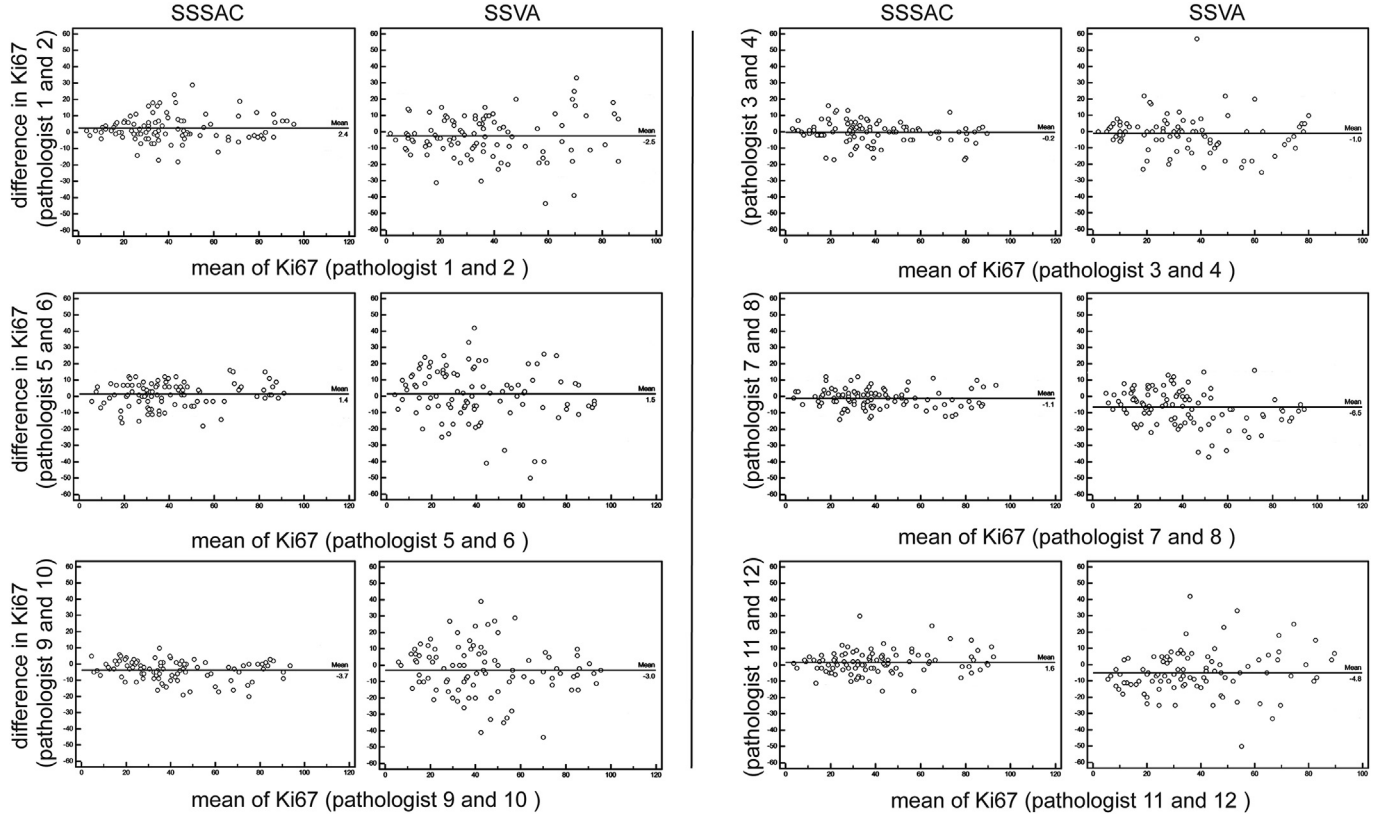


**Fig. 3.** Bland-Altman Plots analysis of Ki67 LI for 100 cases of breast cancer from 6 paired pathologists. The y-axis represents difference in Ki67 LI assessed by two pathologists. The middle solid line represents the average of the differences from the 2 pathologists. Differences from SSSAC were more concentrated to the middle solid line than that from SSVA.

larger than that with SSDIA (negative ranks (SSDIA < SSSAC): N = 80, sum of ranks = 4274.50; positive ranks (SSDIA > SSSAC): N = 17, sum of ranks = 478.50; Z = −6.837, P < 0.001).

### 3.4. Reproducibility of the different groups

According to the number of years that the pathologists were engaged in pathological diagnosis, 41 pathologists were divided into an experienced group (n = 35) and an inexperienced group (n = 6). The ICCs were relatively lower in the inexperienced group than in the experienced group using SSSAC and SSVA. SSSAC showed better reproducibility than SSVA in the two groups (Table 2). The experienced group showed almost perfect correlation using SSSAC (ICC, 0.953 [95% CI: 0.940, 0.965]) and SSVA (ICC, 0.805 [95% CI: 0.760, 0.849]). The inexperienced group showed almost perfect correlation using SSSAC (ICC, 0.903 [95% CI: 0.873, 0.928]) and a substantial correlation using SSVA (ICC, 0.786 [95% CI: 0.730, 0.837]).

To analyze the differences in the Ki67 LI assessment caused by the distribution of Ki67-positive tumor cells, we divided all slides into three groups, the Ki67 homogeneous staining group (n = 16), the centralized Ki67 hot-spot group (n = 33) where hot spots are apparent and concentrated, and the scattered Ki67 hot-spot group (n = 51) where hot spots are noncompact and small. The reproducibility of Ki67 LI assessment in the scattered hot-spot group was obviously lower than that in the homogeneous and centralized hot-spot groups (Table 2). SSVA presented an almost perfect correlation in the homogeneous group (ICC, 0.876 [95% CI: 0.791, 0.945]), a substantial correlation in the centralized hot-spot group (ICC, 0.745 [95% CI: 0.670, 0.820]) and a moderate correlation in the scattered hot-spot group (ICC, 0.648 [95% CI: 0.560, 0.742]). SSSAC showed a perfect correlation in the homogeneous group (ICC, 0.967 [95% CI: 0.938, 0.987]), centralized hot-spot group (ICC, 0.949 [95% CI: 0.922, 0.970]), and scattered hot-spot group (ICC, 0.862 [95% CI: 0.812, 0.906]).

Furtherly, 41 pathologists were divided into two groups according to the location of the laboratory where they worked, provincial hospital laboratory and state hospital laboratory. The ICCs showed no significant difference between the provincial hospital laboratory and state hospital laboratory whether using SSVA or SSSAC (Table 2).

To find out whether histological grade had an effect on the reproducibility of Ki67 assessments using these two methods, ICC of grade 1, 2, 3 breast cancer by 41 pathologists were calculated. We found that SSSAC achieved better ICC in grade 1, 2 and 3 breast cancer, indicating that histological grade had no effect on the reproducibility when SSSAC was used. The ICC of SSVA was higher in grade 2 breast cancer, but was lower in grade 1 and grade 3 breast cancer, indicating that the reproducibility of Ki67 assessments would be affected by tumor grade when SSVA was used (Table 2).

## 4. Discussion

Ki67 plays an indispensable role in the molecular subtyping of breast cancer and personalized treatment. However, the lack of stability in Ki67 LI assessment is a major obstacle for the development of personalized therapies [22] and hinders the confident use of Ki67 LI in clinical decisions. The inconsistency of assessment methods is an important reason that leads to poor reproducibility in Ki67 LI assessment. In this study, we recommended a modified Ki67 assessment, size-set semiautomatic counting (SSSAC), which showed satisfactory interobserver reproducibility. SSSAC has some advantages which overcome the factors that cause poor reproducibility of Ki67 assessment, such as defining only hot spots as the scoring area, restricting the size of the scoring area using frames and identifying tumor cells artificially.

SSSAC defines only Ki67 hot spots as the scoring area. Nonconformity of scoring areas is the main factor that affects the reproducibility of Ki67 assessment. Most of the Ki67 staining slides of breast cancer are heterogeneous, including hot spots, cold spots, periphery areas and areas of intermediate proliferation [23], and Ki67 positive cells in each area are different remarkably. Therefore, the choice of scoring area could cause large differences in Ki67 LI. International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group recommended that, if there are clear hot spots, data from these areas should be included in the overall score [7]. Gudlaugsson E et al. reported that Ki67 in the hot spots is of utmost prognostic significance compared with the average and lowest zones [11]. In our study, we divided all the slides into the homogeneous, centralized hot-spot and scattered hot-spot groups. Both SSSAC and SSVA suggested a better correlation in the centralized hot-spot and homogeneous groups compared with the scattered hot-spot group. The reason may be that ensuring the consistency of the scoring area in the centralized hot-spot and homogeneous groups is easier than that in the scattered hot-spot group. Thus, SSSAC only considers hot spots as scoring areas to overcome the negative influence caused by the nonconformity of the scoring areas.

Our SSSAC restricted the size of the scoring area using frames to guarantee similar numbers of counting cells among observers. The size of the scoring field is another factor that causes poor reproducibility of the Ki67 assessment because different sizes of the scoring field cover different cell numbers [14]. Mikami Y et al. printed photographs of the scoring fields of Ki67 LI to avoid variations in the assessment in various microscopic fields and yielded better concordance than Ki67 LI assessment under a microscope.

The ICC and Bland Altman plot analyses revealed that the reproducibility of SSSAC was better than that of SSVA. SSSAC calculated Ki67 LI by counting each tumor cell using mouse and software in which the counted tumor cells were covered by a crucial marker with the aid of software to avoid repeat counting or missing positive cells that occurred in visual scan and manual counting under a microscope. More than one group demonstrated that counting on real slides affected Ki67 assessment reproducibility. Gudlaugsson E et al. mentioned that the interobserver variation of manual counting was significant [11]. In Mikami Y 'study the manual counting had a moderate correlation (ICC, 0.66 [95% CI 0.52−0.78]) [24]. Varga, Z et al. found that the standard deviations around the mean values were large when Ki67 LI were obtained by counting Ki67-positive nuclei in real slides [18].

Additionally, SSSAC-identified tumor cells could avoid the errors observed using DIA. The Wilcoxon signed-rank test showed that Ki67 LI by SSSAC was obviously higher than that by SSDIA. The cause may be that parameter setting by pathologists used in SSDIA to distinguish Ki67-positive and -negative tumor cells cannot recognize every type of breast cancer cell exactly. For example, Ki67-negative lymphocytes and stroma cells may be recognized as Ki67-negative tumor cells, Ki67-positive tumor cells may not be recognized, and 2 contiguous tumor cells may be recognized as 1 tumor cell type.

Although SSSAC has many advantages mentioned above in improving the interobserver reproducibility for Ki67 assessment, there are some weak spots compared with traditional visual assessment. SSSAC is a little bit time-consuming. The method involves slide scanning, hotspot searching, scoring frame demarcating and semi-automatic cell counting, and would take about 8 min for each case. But in order to provide an accurate and reliable Ki67 value LI for precise treatment of breast cancer patients, this time consumption is worthwhile.

In this study, only NOS invasive ductal carcinomas were employed and other histological types of breast cancer such as tubular carcinoma, cribriform carcinoma, papillary carcinoma, and so on, were not involved. The correlation between hot spots and different histological types, as well as the assessment concordance related to different histotypes should be explored in the future.

Breast cancer subtypes based on four marker surrogate immunohistochemistry panel, including estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67, provide predictive and prognostic information for breast cancer patients [25,26]. Meanwhile, genomic testing has been of vital importance in the prognosis and prognostic of breast cancer, such as 21-gene [27], 70-gene [28], 50-gene [29] and so on.

In conclusion, SSSAC possesses the accuracy of manually identifying tumor cells, convenience of semiautomatic counting and satisfactory interobserver reproducibility. We consider SSSAC to be a promising candidate as a standard method for Ki67 LI assessment in breast cancer as well as in other malignancies.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Ethical approval

All authors declare that the work has been carried out in accordance with The Code of Ethics of the World medical Association (Declaration of Helsinki). The studies have been approved by the ethics committee of Kunming General Hospital of Chengdu Military Command (920th Hospital of Joint Logistics Support Force of PLA) (2015035).

## Declaration of competing interest

There were no conflict of interest relevant to this article.

## Acknowledgements

The authors thank the Department of Pathology at the following hospitals for their participation and contribution: 920th Hospital of Joint Logistics Support Force of PLA; West China Hospital, Sichuan University; Fudan University Shanghai Cancer Center; First Affiliated Hospital of Kunming Medical University; Second Affiliated Hospital of Kunming Medical University; Yunnan Cancer Hospital; First People's Hospital of Yunnan Province; Second People's Hospital of Yunnan Province; Third People's Hospital of Yunnan Province; First People's Hospital of Kunming; Yan' An Hospital of Kunming; Yunnan Provincial Hospital of Traditional Chinese Medicine; Kunming Calmette International Hospital; Fifty-ninth hospital of PLA; People's Hospital of Lvchun County of Yunnan Province, People's Hospital of Xinping Yi and Dai Autonomous County of Yunnan Province; People's Hospital of Wenshan State of Yunnan Province; First People's Hospital of Honghe state of Yunnan Province; People's Hospital of Dehong state of Yunnan Province; First Affiliated Hospital of Dali University of Yunnan Province; First People's Hospital of Chuxiong state of Yunnan Province; People's Hospital of Tengchong City of Yunnan Province; People's Hospital of Luoping county of Yunnan Province; People's Hospital of Lincang City of Yunnan Province.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.breast.2019.12.009.

## References

[1] Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. CELL Cycle analysis of a cell proliferation-associated humannuclear antigen defined by the monoclonal antibody Ki-67. J Immunol 1984;133(4):1710–5.
[2] Gerdes J, Schwab U, Lemke H, Stein H. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. Int J Cancer 1983;31(1):13–20.
[3] Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol 2010;11(2):174–83.
[4] Viale G, Giobbie-Hurder A, Regan MM, et al. Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. J Clin Oncol 2008;26:5569–75. https://doi.org/10.1200/JCO.2008.17.0829.
[5] de Azambuja E, Cardoso F, de Castro Jr G, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Canc 2007;96:1504–13. https://doi.org/10.1038/sj.bjc.6603756.
[6] Sheri A, Dowsett M. Developments in Ki67 and other biomarkers for treatment decision making in breast cancer. Ann Oncol 2012;23(Suppl 10):x219–27. https://doi.org/10.1093/annonc/mds307.
[7] Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group. J Natl Cancer Inst 2011;103:1656–64. https://doi.org/10.1093/jnci/djr393.
[8] Kim JY, Jeong HS, Chung T, Kim M, Lee JH, Jung WH, Koo JS. The value of phosphohistone H3 as a proliferation marker for evaluating invasive breast cancers: a comparative study with Ki67. Oncotarget 2017;8(39):65064–76.
[9] Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013;105:1897–906. https://doi.org/10.1093/jnci/djt306.
[10] Polley MY, Leung SC, Gao D, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015;28:778–86. https://doi.org/10.1038/modpathol.2015.38.
[11] Gudlaugsson E, Skaland I, Janssen EA, 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;61:1134–44. https://doi.org/10.1111/j.1365-2559.2012.04329.x.
[12] Denkert C, Budczies J, von Minckwitz G, Wienert S, Loibl S, Klauschen F. Strategies for developing Ki67 as a useful biomarker in breast cancer. Breast 2015;24:S67–72. https://doi.org/10.1016/j.breast.2015.07.017.
[13] Stalhammar G, Fuentes Martinez N, Lippert M, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. Mod Pathol 2016;29:318–29. https://doi.org/10.1038/modpathol.2016.34.
[14] Christgen M, von Ahsen S, Christgen H, Langer F, Kreipe H. The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer. Hum Pathol 2015;46:1341–9. https://doi.org/10.1016/j.humpath.2015.05.016.
[15] R N. A revolution postponed indefinitely. WHO classification of tumors of the breast 2012: the main changes compared to the 3rd edition. Cesk Patol 2003;51:3. 2015.
[16] Leung SCY, Nielsen TO, Zabaglo LA, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. Histopathology 2019;75:225–35. https://doi.org/10.1111/his.13880.
[17] Leung SCY, Nielsen TO, Zabaglo L, et al. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. NPJ Breast Cancer 2016;2:16014. https://doi.org/10.1038/npjbcancer.2016.14.
[18] Varga Z, Diebold J, Dommann-Scherrer C, et al. 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 2012;7:e37379. https://doi.org/10.1371/journal.pone.0037379.
[19] Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 1973;1973(33):613–9.
[20] Richard Landis J, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
[21] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307–10.
[22] Laurinavicius A, Laurinaviciene A, Dasevicius D, et al. Digital image analysis in pathology: benefits and obligation. Anal Cell Pathol 2012;35:75–8. https://doi.org/10.3233/ACP-2011-0033.
[23] Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. Breast Canc Res Treat 2016;159(2):257–63. https://doi.org/10.1007/s10549-016-3950-5.
[24] Mikami Y, Ueno T, Yoshimura K, et al. Interobserver concordance of Ki67 labeling index in breast cancer: Japan Breast Cancer Research Group Ki67 ring study. Cancer Sci 2013;104:1539–43. https://doi.org/10.1111/cas.12245.
[25] Goldhirsch A, Winer EP, Coates AS, 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. Ann Oncol 2013;24:2206–23. https://doi.org/10.1093/annonc/mdt303.

[26] Cuzick J, Dowsett M, Pineda S, 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. J Clin Oncol 2011;29:4273–8. https://doi.org/10.1200/JCO.2010.31.2835.

[27] Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. J Clin Oncol 2010;28:1829–34. https://doi.org/10.1200/JCO.2009.24.4798.

[28] Bueno-de-Mesquita JM, van Harten WH, Retel VP, et al. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). Lancet Oncol 2007;8:1079–87. https://doi.org/10.1016/s1470-2045(07)70346-7.

[29] Sestak I, Cuzick J, Dowsett M, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. J Clin Oncol 2015;33:916–22. https://doi.org/10.1200/JCO.2014.55.6894.