The clinicopathological potential of Ki67 labeling index in breast cancer

Synopsis of PhD thesis

Balázs Ács M.D.

Doctoral School of Pathological Sciences
Semmelweis University

Supervisor: Attila Marcell Szász M.D., Ph.D., consultant pathologist

Official reviewers:
Lilla Reiniger M.D., Ph.D., assistant professor
Ilona Kovács M.D., Ph.D., consultant pathologist, head of division

Head of the Final Examination Committee:
Péter Nagy M.D., D.Sc., professor

Members of the Final Examination Committee:
Katalin Dezső M.D., Ph.D., assistant professor
László Landherr M.D., Ph.D., consultant oncologist, head of division

Budapest
2017

INTRODUCTION

Ki67 Labeling Index (Ki67 LI) is currently one of the most promising yet controversial biomarker in breast cancer. It has been long acknowledged - and more recently several studies have demonstrated - that the immunohistochemical (IHC) detection of the Ki67 positive tumor cells provides important prognostic information in breast cancer. The St. Gallen Consensus Conference in 2015 recommended the use of Ki67 LI to distinguish between HER2 negative luminal B-like and luminal A-like breast carcinomas. However, the International Ki67 in Breast Cancer Working Group is more cautious about the recommendation for use of Ki67 in daily practice. The European Society for Medical Oncology (ESMO) Clinical Practice Guidelines suggests that Ki67 LI may provide useful information, provided the assay can be standardized. The American Society of Clinical Oncology (ASCO) did not recommend the use of Ki67 LI for prognosis in newly diagnosed breast cancer patients because of lack of reproducibility across laboratories.

In addition to ongoing debate on its prognostic utility, Ki67 LI has also been investigated as a potential predictive marker for response to therapy. In a recent research involving 506 breast cancer patients, Ki67 LI did not represent an independent predictive potential for neoadjuvant chemotherapy. In contrast to this, the systematic review by Luporsi et al. has determined a level of evidence of II-B for Ki67 LI regarding neoadjuvant treatment response. Despite the promise of Ki67 LI as a

prognostic and/or predictive tool, controversy exists regarding its applied methodology in practice. Therefore, there is an urgent need for reproducible methodology and consistent scoring methods of Ki67 LI. To overcome this struggle, the International Ki67 in Breast Cancer Working Group has introduced a recommendation for the application of Ki67 IHC in daily practice. According to this, parameters that predominantly influence the Ki67 IHC results include pre-analytical (type of biopsy, tissue handling), analytical (IHC protocol), interpretation and scoring, and data analysis steps. Serious efforts have been made to improve the pre-analytical and analytical validity of Ki67 IHC. However, little emphasis had been put so far on a very evident technical question, namely, are all commercially available Ki67 antibodies detecting the same amount of proliferating tumor cells in each case? Can we use the different antibodies interchangeably? Most published studies concluded that there are indeed differences between the protein expression levels detected by different Ki67 antibodies; however, to the best of my knowledge, the different results were never linked to the prognosis.

Difficulties in evaluating immunoreactions can also be responsible for discrepancies of Ki67 scoring reproducibility. Ki67 LI values are usually defined as the percentage of positive tumor cell nuclei, counted in 3-10 high-power fields by testing at least 500-1000 tumor cells. Another method is to estimate the mean Ki67 LI in the entire lesion. Both methods are monotonous, time-consuming and exhausting with a

chance of leading to controversial results and inaccurate reproducibility. Although the counting method has been recommended by the International Ki67 in Breast Cancer Working Group, other studies have demonstrated the counting method is not superior to visual estimation. Although, recommendations published in 2011 provide a suitable landmark to improve pre-analytical and analytical validity, related protocols still show high variety and poor reproducibility linked with the context of different sampling, fixation, antigen retrieval, staining and scoring methods.

Rapid development of digital microscopy by now allows fast digitalization of histological slides at high-resolution, which can firmly support education, research and diagnostics in pathology. The emergence of digital image analysis (DIA) platforms improved the capacity, precision and reproducibility of in situ biomarker evaluation. However, these features alone may not be enough for diagnostic accuracy, which must be based on histological pattern recognition as the most relevant requirement of precise sample selection and assessment of immunohistochemical reactions. DIA platforms are able to assess Ki67 LI, however it has not been clarified yet, whether their results can meet the requirements of the daily diagnostic practice and reduce variability of Ki67 scoring.

OBJECTIVES

Three aspects of clinical validity of Ki67 LI are investigated as follows: i) The comparison of different Ki67 antibodies used in daily practice. ii) The reproducibility between pathologists evaluating Ki67 LI and the potential of DIA in Ki67 scoring. iii) The role of Ki67 in neoadjuvant setting. Thus, we aimed to:

1, Compare the semi-quantitatively defined Ki67 LI of five commercially available Ki67 IHC antibodies in a consecutive breast cancer patient population.

2, Correlate the prognosis prediction potential of each Ki67 antibodies with that of conventional clinicopathological factors in univariate and multivariate analyses.

3, Investigate the reproducibility of Ki67 LI among three pathologists, based on their conventional visual estimation.

4, Test the agreement of semi-quantitative and DIA Ki67 scoring.

5, Determine and compare the outcome prediction potential of each semi-quantitative and DIA assessments with that of conventional clinicopathological factors.

6, Find optimal cut-off values for Ki67 LI in neoadjuvant setting that best correlates with response rates to neoadjuvant therapy.

7, Investigate the association between Ki67 LI, subtype and pathological response.

8, Investigate the prognostic potential of Ki67 LI in neoadjuvant setting with multivariate analysis.

METHODS

Two distinct breast cancer patient cohorts were enrolled in the investigations encompassing 498 patients totally without any overlap: 1) 378 consecutive breast cancer cases from the Buda MÁV Hospital Pathology Unit, Budapest, Hungary diagnosed between 1999 and 2002 (ethical approval: TUKEB, #7-1/20) with 99.80 months median follow up (disease-free survival, DFS). 2) 120 patients diagnosed with invasive breast cancer and treated with neoadjuvant chemotherapy (NAC) at Semmelweis University, Budapest, Hungary between 2002 and 2013 were retrospectively recruited (ethical approval: TUKEB 120/2013). The median follow up time for overall survival (OS) and distant metastases-free survival (DMFS) was 60.5 and 59 months, respectively. Degree of response to NAC was categorized according to Pinder et al. (2007) using the definitions as follows: pathologic complete response (pCR), partial response to therapy (pPR), no response to therapy (pNR). Regarding the definition of surrogate molecular subtypes of breast cancer, we referred to the St. Gallen recommendations from 2013 that include five categories (luminal A, luminal B/HER2-, luminal B/HER2+, HER2+ and triple negative).

Tissue microarrays (TMA) were built from 10% neutrally buffered FFPE representative tissue blocks of the 378 consecutive cases. Duplicate cores (each 2 mm in diameter) were punched (TMA Master, 3DHISTECH Ltd., Budapest, Hungary) from each case, resulting 10 TMA blocks. Regarding the neoadjuvant cohort involving 120 cases,

the pre-treatment core biopsy specimens and in case of non pCR, the surgical specimens were also investigated.

Paraffin sections of 3 μm thickness were cut from the TMA blocks for IHC. The following five antibodies were used for IHC detection of Ki67 on TMA blocks: SP6 (Histopathology), 30-9 (Ventana), N1574-poly (DAKO), B56 (Histopathology), MIB1 (Immunotech). Furthermore, Ki67-MIB1 was investigated with immunofluorescent labeled (MIB1-IF) antibody (IR 626 DAKO) as well. The IHC reactions were performed in an automated immunostainer (Ventana Benchmark XT, Roche, Basel, Switzerland) according to the manufacturer's protocol. To detect Ki67 in core biopsy and surgical specimens of the neoadjuvant breast cancer cohort MIB1 antibody with chromogen detection was used with the same protocol.

Semi-quantitative (SQ) evaluation of Ki67 IHC of 378 consecutive cases was performed on digital slides using the TMA Module software on the PannoramicViewer platform as follows: Ki67 LI was defined as the percentage of positive tumor cell nuclei, estimated on average in 3-10 high-power fields, in each core. Any nuclear positivity was considered in a range of 100–500 cells, depending on the cellularity of the TMA cores. Duplicate cores were evaluated separately and their mean Ki67 LI was finally analyzed.

During the comparison of five Ki67 antibodies, the IHC reactions were evaluated by two pathologists independently and if any discrepancy

occurred, the inconsistent cases were reassessed and a consensus Ki67 LI score was given.

When the reproducibility was investigated between observers, the IHC reactions of MIB1 antibody were evaluated by three pathologists (SQ-1, SQ-2, SQ-3) independently. The three pathologists have considerable but different level of experience in Ki67 scoring of breast cancer. SQ1 is the youngest with a pathology specialist status for a year only. SQ-2 and SQ-3 are consultant pathologists with substantial experience in diagnostic practice and special focus on breast pathology. Regarding the neoadjuvant cohort, the Ki67 IHC reactions were evaluated by two pathologists independently and if any discrepancy occurred, the inconsistent cases were reassessed and a consensus Ki67 LI score was given.

TMA slides were digitized with Pannoramic Flash II slide scanner using x20 objective (NA=0.83). DIA was performed on the IHC reactions of MIB1 antibody using the PatternQuant (PQ) software enabling automated tissue pattern recognition by separating epithelial elements from stroma. All digital hardware and software tools were from 3DHISTECH Ltd. (Budapest, Hungary). Designation of training tissue patterns to be recognized and the calibration were done in co-operation by a pathologist and an IT expert to achieve the best recognition pattern (achieved at a PQ training magnification of 1.5x; a gamma level of 1; dilution of 3; a contour of 0). So, as the detection and quantification of tumor cell nuclei using NuclearQuant (NQ) at the following settings:

Blur: 15; Radius minimum: 1.5; Radius maximum: 8; Area min: 15; Intensity minimum: 30; Contrast minimum 30. Based on these settings of PQ and NQ, automated Ki67 evaluation was performed on each core (DIA-1 analysis). In the other DIA test, automated annotations were assessed by pathologists on each core, and when it was necessary, DIA settings were adjusted independently to exclude artifacts, underestimation or overestimation of positive/negative cells and false detections (DIA-2 analysis).

Degree of agreement among different antibodies detecting Ki67 and the reproducibility between Ki67 evaluations was evaluated by using intra-class correlation coefficient (ICC), concordance correlation coefficient (CCC), Cohen's kappa and Bland-Altman plot (MedCalc 13.3.3.0, Ostend, Belgium). To assess statistical differences between each antibody and between observers, Wilcoxon signed-rank and McNemar tests were applied, since our data were not normally-distributed, even after log-transformation (Shapiro-Wilk and Kolmogorov-Smirnov tests). The optimal cut-off value for Ki67 percentage to discriminate response to treatment was assessed by receiver operating characteristic (ROC) curve analysis. To identify the optimal Ki67 threshold for NAC, only pCR and pNR cases were involved in ROC analyses, because pPR status is considered as a soft endpoint. Kaplan-Meier analysis supported with log-rank test and multivariate Cox-regression was executed to assess prognostic potential (SPSS 22 software, IBM, Armonk, USA). In all statistical analysis, the level of significance was set at $p < 0.05$.

RESULTS

The validity of five Ki67 antibodies

The Ki67 LI scores of the 5 antibodies showed a moderate agreement. Highest concordance was observed between MIB1 and poly, 30-9 and poly, MIB1 and B56, 30-9 and SP6 as well as between MIB1 and 30-9. Conversely, lowest agreement was found between SP6 and B56 as well as between SP6 and MIB1-IF. Significant bias was observed in all comparisons except MIB1 vs. MIB1-IF and the range of agreement was also wide. Furthermore, a systematic error was found between all the antibodies except between MIB1 and poly. For prognosis, all the Ki67 antibodies but the IF detection of MIB1 could perform statistically significant splitting of our cohort into 2 patients' groups with distinct DFS at 20% threshold. At 30% cut-off point, Ki67 LI of MIB1, SP6, 30-9 and poly could distinguish good and unfavorable prognosis patients' cohorts. Meanwhile B56 and MIB1-IF did not represent any statistically significant prognosis predictor potential at 30% threshold. We had also investigated the utility of each Ki67 antibodies as potential independent predictors of DFS adjusted by age, IHC subtypes, lymph node and T status, histological grade, mitotic index, vascular invasion as well as necrosis at 20% and 30% thresholds. At 20% cut-off score, Ki67 LI of poly and lymph node status were significantly linked to DFS However, at 30% threshold, only lymph node status represented an independent association with survival.

The reproducibility between different Ki67 evaluations

Since MIB1 is the most widely used antibody to detect Ki67 LI and showed the highest concordance and agreement with the poly antibody, it was used in the further investigations.

The 3 SQ Ki67 LI assessments showed a very good consistency concerning the relative difference between cases. The best interobserver variability was found between SQ-2 and SQ-3 while SQ-1 showed poor concordance with SQ-2 and SQ-3. Significant bias was observed in all comparisons. The lowest bias and the narrowest range of agreement were found between SQ-2 and SQ-3 without a systematic error. Upon dichotomizing Ki67 LI values at 14% and 20% thresholds, SQ-1 still differed considerably from SQ-2 and SQ-3 with a moderate agreement. However, no significant difference, and substantial agreement were found between SQ-2 and SQ-3, at these thresholds.For the comparison with DIA assessments, a reference SQ Ki67 LI value was generated (SQ-RV) as the mean of SQ-2 and SQ-3, since SQ-1 differed considerably from those. SQ-RV and automated DIA-1 differed and showed moderate concordance. SQ-RV and adjustable DIA-2 showed no significant difference, and represented a substantial concordance. Significant difference but substantial concordance was found when DIA-1 was compared to DIA-2. A significant bias and proportional error was found between SQ-RV and DIA-1 values, which was not seen between SQ-RV and DIA-2 values and the range of agreement was also

superior in the latter case. For prognosis, all Ki67 evaluations but SQ-3 could perform statistically significant splitting of our cohort into 2 patients' group with distinct DFS at 14% threshold. At 20% cut-off point, Ki67 evaluations of DIA-2, SQ-2 andSQ-3 could sort patients into good and unfavorable prognostic groups, while SQ-1 and DIA-1 did not. Ki67 LI assessments were also tested as potential independent predictors of DFS adjusted by age, IHC subtypes, lymph node and T status, histological grade, mitotic index, vascular invasion as well as necrosis. At 14% cut-off, no Ki67 LI evaluation but only lymph node status showed independent association with DFS. However, at 20% threshold, both lymph node status and SQ-2 were significantly linked to DFS.

The role of Ki67 in neoadjuvant setting

The optimal Ki67 LI cut-off value was 20% for distinguishing pCR from pNR patient cases. Pathological response and Ki67 LI at investigated thresholds represented a significant association. The proportion of Ki67 LI low cases among non-responders was significantly higher compared to pPR and pCR cases. The distribution of subtypes showed a significant difference in pathological response groups. Most of the TNBC cases were represented in pCR group, while luminal A cases mainly occurred in pPR and pNR groups. The Ki67 expression at any investigated cut-off points and subtypes also represented a significant correlation. Luminal A subtype showed low

Ki67 LI, while TNBC and HER2+ cases mostly had high Ki67 LI. The association between Ki67 LI, subtype and pathological response was also investigated without luminal A cases, because NAC is not generally recommended in this subtype due to the high rate of pNR in contrast with the favorable prognosis. Excluding luminal A cases, Ki67 LI at any thresholds and pathological response did not show any significant association. Furthermore, Ki67 LI at any investigated cut-off points also did not represent any significant linkage with subtypes. In contrast to this, subtypes were significantly linked to the pathological response groups. The clear majority of luminal B cases were in pPR and pNR groups, while TNBC cases mostly occurred in pCR subgroup. Regarding OS, Ki67 LI at 15% and at 20% threshold failed, but Ki67 LI at 30% cut-off value, furthermore subtype as well as pathological response were suitable to separate patients into good and unfavorable prognosis cohorts. When luminal A cases were excluded, neither Ki67 LI at any cut-off points nor subtype not even pathological response were suitable to perform statistically significant splitting of our cohort into 2 patients' group with different OS. We also investigated the utility of Ki67 LI at 15%, 20% and 30% thresholds as potential independent predictor of OS adjusted by age, pathological response, hormone receptor status, subtypes, histological grade, lymph node, cT and pT status. Ki67 LI at 30% threshold and subtype were independently linked to OS. Without luminal A cases, Ki67 LI at 30% cut-off point and subtype represented also an independent association with OS.

CONCLUSIONS

In surgical pathology practice, the selection and then the validation of Ki67 antibody requires great caution. Our results suggest that, as MIB1, poly, 30-9 antibodies showed the highest performance, they are suitable to detect Ki67 expression in the daily practice. We believe that this study provides a partial explanation to the various suggested Ki67 LI cut-off values in different published series of breast cancer cases.

The pathologists' experience is essential to control and adjust DIA and to avoid false detections. We also demonstrate that the adjustable DIA can be a feasible and reproducible tool to evaluate Ki67 LI in breast cancer which may support standardization efforts.

Neoadjuvant chemotherapy is more efficient in tumors presenting at least 20% Ki67 LI. A cut-off value of 20% distinguished pCR from pNR cases. Increased Ki67 LI was linked to worse OS, meaning that at least in some subgroups higher Ki67 expression is related to increased response to NAC and is also associated with worse prognosis. Additionally, our data also suggest that if a tumor is non-responder to NAC, increased Ki67 LI is a poor prognostic marker. Moreover, we provide further evidence that Ki67 LI is a significant and independent prognostic marker in breast cancer. Thus, we can conclude that Ki67 has potential utility in the clinical management of breast cancer. However, we can also state that Ki67 LI in itself is not suitable to decide whether a breast cancer patient should be treated with NAC or not.

PUBLICATIONS RELATED TO THE PHD THESIS

1. Acs B, Madaras L, Kovacs A, Micsik T, Tokes AM, Gyorffy B, Kulka J, Szasz AM - **Reproducibility and Prognostic Potential of Ki-67 Proliferation Index when Comparing Digital-Image Analysis with Standard Semi-Quantitative Evaluation in Breast Cancer. –** Pathology & Oncology Research. 2017 Apr 11. doi: 10.1007/s12253-017-0220-8 - IF: 1.736, http://dx.doi.org/10.1007/s12253-017-0220-8

2. Acs B, Zambo V, Vizkeleti L, Szasz AM, Madaras L, Szentmartoni Gy, Tokes T, Molnar BA, Molnar IA, Vari-Kakas S, Kulka J, Tokes AM - **Ki-67 as a controversial predictive and prognostic marker in breast cancer patients treated with neoadjuvant chemotherapy.** – Diagnostic Pathology. 2017 Feb 21. doi: 10.1186/s13000-017-0608-5 – IF: 2.087, http://dx.doi.org/10.1186/s13000-017-0608-5

3. Acs B, Kulka J, Kovacs A, Teleki I, Tokes AM, Meczker A, Gyorffy B, Madaras L, Krenacs T, Szasz AM - **Comparison of five Ki67 antibodies regarding reproducibility and capacity to predict prognosis in breast cancer: Does the antibody matter?** – Human Pathology. 2017 Feb 8. pii: S0046-8177(17)30032-1. doi: 10.1016/j.humpath.2017.01.011 – IF: 3.014, http://dx.doi.org/10.1016/j.humpath.2017.01.011

PUBLICATIONS NOT RELATED TO THE PHD THESIS

1. Acs B, Madaras L, Tokes AM, Kovacs A, Kovacs E, Ozsvari-Vidakovich M, Karaszi A, Birtalan E, Dank M, Szasz AM, Kulka J - **PD-1, PD-L1 and CTLA-4 in pregnancy-related - and in early-onset breast cancer: A comparative study.** - Breast. 2017 Oct;35:69-77. doi: 10.1016/j.breast.2017.06.013 – IF: 2.801, http://dx.doi.org/10.1016/j.breast.2017.06.013

2. Molnar IA, Molnar BA, Vizkeleti L, Fekete K, Tamas J, Deak P, Moldvay J, Vari-Kakas S, Szundi Cs, Szasz AM, Acs B, Kulka J, Tokes AM - **Breast carcinoma subtypes show different patterns of metastatic behavior.** - VirchowsArchiv. 2017 Jan 19. doi: 10.1007/s00428-017-2065-7.– IF: 2.848, http://dx.doi.org/10.1007/s00428-017-2065-7

3. Santarpia L, Bottai G, Raschioni C, Szekely B, Tommaso LD, Szasz AM, Losurdo A, Gyorffy B, Acs B, Torrisi R, Karachaliou N, Tokes T, Caruso M, Kulka J, Roncalli M, Santoro A, Mantovani A, Rosell R, Reis-Filho J. - **AXL associated tumor inflammation as a poor prognostic signature in chemotherapy-treated triple-negative breast cancer patients. –** npj Breast Cancer.2, Article number: 16033 (2016) doi:10.1038/npjbcancer.2016.33. - IF: 0.000, http://dx.doi.org/10.1038/npjbcancer.2016.33

4. Agoston EI, Micsik T, Acs B, Fekete K, Hahn O, Baranyai Zs, Dede K, BodokyGy, Bursics A, Kulka J, Krenacs T, Gyorffy B, Harsanyi L, Szasz AM. – **In depth evaluation of the prognostic and predictive utility of PTEN immunohistochemistry in colorectal carcinomas: performance of three antibodies with emphasis on intracellular and intratumoral heterogeneity**. - Diagnostic Pathology. 2016 Jul 8;11(1):61. doi: 10.1186/s13000-016-0508-0. - IF: 2.087,
http://dx.doi.org/10.1186/s13000-016-0508-0

5. Acs B, Szekely N, Szasz AM, Lotz G, Szekely T, Istok R, Szekely E, Madaras L, Kulka J, Jaray B. - **Reliability of immunocytochemistry and fluorescence in situ hybridization on fine needle aspiration cytology samples of breast cancers: a comparative study.** – Diagnostic Cytopathology. 2016 March, DOI: 10.1002/dc.23463. - IF: 1.161,
http://dx.doi.org/10.1002/dc.23463

6. Acs B, Szarvas T, Szekely N, Nyirady P, Szasz AM - **Current State of ERG as Biomarker in Prostatic Adenocarcinoma.** – Current Cancer Drug Targets. 2015;15(8):643-51. - IF: 3.707, http://dx.doi.org/10.2174/1568009615081510011008 29

7. Cserni G, Bori R, Sejben I, Agoston EI, Acs B, Szasz AM - **The Petersen prognostic index revisited in Dukes B colon cancer – inter-institutional differences.** - Pathology Research and Practice. 2015 September, DOI: 0.1016/j.prp.2015.08.006. – IF: 1.543, http://dx.doi.org/10.1016/j.prp.2015.08.006

8. Kulka J, Tokes AM, Madaras L, Kovacs A, Acs B, Illyes I, Kiss O, Szekely B, Lotz G, Szasz AM – [**Clinico-pathologically focused breast cancer research.]** – Hungarian Oncology. 2015;59(4):286-91.– IF: 0.000, http://www.ncbi.nlm.nih.gov/pubmed/26665188

9. Acs B, Szasz AM, Kulka J, Harsanyi L, Zarand A – [**Is radicality enough? Transanal endoscopic microsurgery for the treatment of rectal neoplasia – clinicopathological viewpoint.]** – Hungarian Surgery. 2014; 67(6):329–333. – IF: 0.000, http://dx.doi.org/10.1556/MaSeb.67.2014.6.2

10. Szasz AM, Acs B, Agoston EI, SztupinszkiZs, Tokes AM, Szittya L, Szekely B, Szendroi M, Li Q, Harsanyi L, Timar J, Szallasi Z, Swanton C, Gyorffy B, Kulka J – [**Simplified, low-cost gene expression profiling for the prediction of outcome in breast cancer based on routine histologic specimens.]** – Hungarian Weekly Journal of Medicine. 2015; 154(16):627–632. – IF: 0.000, http://dx.doi.org/10.1556/OH.2013.29590