Abstract. Background: Ki-67 is a widely used proliferation marker reflecting prognosis in various tumors. However, visual assessment and scoring of Ki-67 suffers from marked inter-observer and intra-observer variability. We aimed to assess the concordance of manual counting and automated image-analytic scoring methods for Ki-67 in synovial sarcoma. Patients and Methods: Tissue microarrays from 34 patients with synovial sarcoma were immunostained for Ki-67 and scored both visually and with 3DHistech QuantCenter. Results: The automated assessment of Ki-67 expression was in good agreement with the visually counted Ki-67 \( (r_{Pearson}=0.96, \ p<0.001) \). In a Cox regression model automated [hazard ratio (HR)=1.047, \( p=0.024 \)], but not visual (HR=1.063, \( p=0.053 \)) assessment method associated high Ki-67 scores with worse overall survival. Conclusion: The automated Ki-67 assessment method appears to be comparable to the visual method in synovial sarcoma and had a significant association to overall survival.

Synovial sarcoma is a high-grade soft-tissue malignancy accounting for 5-10% of all soft tissue sarcomas (1). The two main histological subtypes show either both mesenchymal and epithelial differentiation (biphasic subtype) or mesenchymal alone, spindle cell-like differentiation (monophasic subtype). Nearly all synovial sarcomas carry a specific t(X;18) chromosomal translocation. Besides this cytogenetic abnormality, synovial sarcoma seems to have a relatively stable genome and a low mutational burden (2).

Ki-67 is a proliferation marker protein, which is expressed through the cell cycle excluding the G0 phase (3). Ki-67 is a widely accepted proliferation marker and is prognostic in various tumors, including soft tissue sarcomas (4). However, visual assessment of Ki-67 scoring has been shown to suffer from a marked inter-observer and intra-observer variability (5, 6). Automated image analysis techniques may improve reproducibility of Ki-67 scoring and be beneficial in handling the increased workload in pathology laboratories (7-11).

The aim of the present study was to assess concordance of visual and automated scoring methods for Ki-67, and to study the prognostic value of both visual and automated method in synovial sarcoma.

Patients and Methods

Patient samples. The use of clinicopathological data and formalin-fixed, paraffin-embedded tissues in this research was approved by the Joint Ethics Committee of Helsinki University Central Hospital and by the Ministry of Health and Social Affairs. The tissues were derived from 34 synovial sarcoma patients referred for primary or locally recurred local disease of the extremities or trunk wall and treated with curative intention by the Soft Tissue Sarcoma Group at Helsinki University Hospital, Helsinki, Finland during 1987-2014. Cytogenetic testing was performed in 15 patients and 13 of them had a diagnostic translocation. In the other 21 cases, the diagnosis was based on morphology and immunohistochemical profile.

Treatment was carried out according to the group’s prospective treatment protocol set up in 1987. In short, surgery with large negative margin is the treatment of choice. Adjuvant radiation
therapy is offered if the definite margin is intraluminal or marginal, defined as less than 2.5 cm according to our treatment protocol. Since 1998, patients with high-risk disease are offered adjuvant combination chemotherapy. High-risk patients were defined according to tumor size (≥5 cm), presence of necrosis or vascular invasion. Adjuvant chemotherapy was recommended if at least two of these prognostic factors were present. All patients have a regular follow-up for 10 years.

TMA construction and immunohistochemistry. In-house constructed tissue microarrays (TMA) of soft tissue sarcoma were used for retrospective analysis in the present study, including 34 specimens from synovial sarcoma patients. In short, a tumor cell-rich area of each specimen was identified by a pathologist (MS). Necrosis and vascular structures were avoided. TMAs were constructed in quadruplicate (or in cases of scarce material, 1-3 cores; core diameter was 1.0 mm) in order to account for intra-tumor heterogeneity. A few tissue spots were lost due to tissue processing or regarded as non-representative because of imperfect scanning, leaving 15 subjects with quadruplicate, 4 with triplicate, 13 with duplicate images and 2 with one TMA core image (100 cores in total).

The tissue sections were cut into 5 μm thick and deparaffinized in xylene for 10 min. Antigen retrieval was performed with ULTRA Cell Conditioning CC1 (Roche 950-224) (Roche Diagnostics GmbH, Mannheim, Germany) for 64 min at 98°C. Dako anti-Ki-67 (monoclonal mouse anti-human, Ki-67 antigen, clone MIB-1; DAKO, Glostrup, Denmark) was used at a dilution of 1:100 for 32 min at 36°C for immunohistochemistry on a Ventana Benchmark Ultra instrument (Roche, Tucson, AZ, USA) according to the manufacturer’s instructions, with the UltraView Universal DAB Detection Kit (Roche 760-500).

Slide scanning and scoring. Stained TMA slides were scanned using the Pannoramic 250® digital scanner (3DHistech, Budapest, Hungary). Visualization and automated cell counts were accomplished using the 3DHistech QuantCenter software. The individual cores were annotated, and cells were counted visually on a computer monitor. Each core had two annotations, one circle-shaped (0.801 mm²) and one rectangular-shaped (0.150 mm²). The rectangular annotation was added to the circular annotation in order to enable visual and automated counting of the identical area with a reasonable workload. The size of the rectangular-shaped annotation was approximately the same as one high power-field (HPF). The circle-shaped annotation included as many tumor cells as possible but excluded the periphery of the sample which often showed artifacts or faint stain due to tissue processing. All tumor cells in the rectangular-shaped annotations were counted both visually and with digital image analysis (DIA), marked as positive or negative and the percentage of positive tumor cells was calculated. The cells in circle-shaped annotations were counted only by the automated method. Ki-67 positivity indexes determined from the rectangular-shaped annotations by the visual and automated assessment methods were compared. Also, the positivity indexes determined from the rectangular- and circle-shaped annotations by the automated method were compared. Nuclei with weak staining were regarded as sufficient for Ki-67 positivity, as proposed by several authors (12, 13). Annotations are exemplified in Figure 1.

Automated image analysis. For the automated analysis of Ki-67 positivity index, digitalized TMA slides were accessed via the CaseViewer system and evaluated using the nuclear scoring algorithm of the 3DHistech QuantCenter software. The algorithm quantifies nuclear staining within an annotation chosen manually and derives a positivity index for each annotation. Color intensities for positive (chromogen) and negative (counterstain) nuclei can be defined by choosing a color sample from the relevant area of the slide image. The observer can also adjust the upper and lower limits of the radius of a nucleus and define the smallest acceptable area for analysis, leaving nuclei out of range to be rejected. This enables the algorithm to eliminate, for instance, inflammatory cells which are much smaller than tumor cells in the case of synovial sarcoma. Nevertheless, there might be some inaccuracy in the selection process, for example, multiple small nuclei overlapping with each other may be counted as a single large nucleus. This kind of errors are possible to correct with the software’s edit shape-tool, but this increases the manual workload considerably. The measurement parameters chosen and set for this study were used for the entire patient cohort and were not adjusted for individual samples.

Statistical analysis. For the survival analysis, only the rectangular-shaped annotation with the highest Ki-67 positivity index (“hot spots”) were assessed. All the rectangular-shaped annotations were included in the comparison of the visual and automated assessment of Ki-67 proliferative activity. The correlation was analyzed with Pearson’s r. Furthermore, Bland-Altman scatterplots were used as a graphical means to assess the agreement between the two analysis methods. There were no treatment-related deaths or deaths not related to sarcoma. Time-to-event outcome parameters [overall survival (OS) and metastasis-free survival (MFS)] were estimated using the Kaplan-Meier method. Continuous Ki-67 values were divided into tertiles for visualization in Kaplan-Meier plots. Calculation of hazard ratios (HR), including 95% confidence intervals (CI) and tests of significance were accomplished using a Cox regression hazards model with proliferation as a continuous variable. The Chi-square test (χ²-value) was used for intergroup comparisons of categorical data. All statistical tests were two-sided and the level of significance was set at 0.05. IBM SPSS Statistics for Windows, Version 25.0 (IBM Corporation, Armonk, NY, USA) was used for all analyses.

Results

The mean age of the patients was 41.7 years (median 40.3, range=17-74 years). Out of 34 patients, 19 (56%) were female. The most common location was lower limb (22 patients), and the median tumor size was 5 cm (range=2-16.5 cm). The median follow-up was 5.6 years (range=1.3-19.5 years) for survivals. Seventeen patients developed metastases, three of them are alive without disease.

Automated Ki-67 expression analysis. A total number of 200 annotations were analyzed from 100 TMA spot images. The mean positivity index for Ki-67 in rectangular-shaped annotations (mean number of cells: 1055), was 9.31% (median: 5.51%, range=0-66.12%). In circle-shaped annotations (mean number of cells: 4,677), the mean positivity index was in good concordance with the mean positivity index calculated from the smaller rectangular-shaped areas, being equal to 8.28% (median: 5.16%, range=0.03-66.22%). Patients were divided
into equal-sized tertile groups according to the automatically assessed Ki-67 expression levels (low<3.2%, moderate 3.2-14.7%, and high>14.7%).

**Visual Ki-67 expression analysis.** Visually assessed mean positivity index for Ki-67 in rectangular-shaped annotations was equal to 7.16% (median: 3.94, range=0.11-42.13%). The mean number of visually counted cells was 1461. For the survival analysis, patients were classified into tertiles according to the visually assessed Ki-67 expression levels (low<2.5%, moderate 2.5-11.3%, high>11.3%).

**Comparison between visual and automated Ki-67 scores.** DIA detected less nuclei than manual counting (difference of 19-1188 nuclei, mean of 406). The number of Ki-67 positive nuclei differed only slightly between the two methods (mean: 9, median: 6, range= –49 to 105) whereas the negatively stained nuclei showed considerably larger differences (~107 to 1,137, mean: 397, median: 340). Specimens with high cell density displayed a larger range in negatively stained nuclei, especially if there was overlapping with adjacent nuclei. A Bland–Altman plot showing agreement between the visual and automated methods is displayed in Figure 2. The automated assessment of Ki-67 expression was strongly correlated with the visually counted Ki-67 (r_{Pearson}=0.96).

**Cases with larger Ki-67 discrepancy.** Eight out of a hundred annotations (8%) showed a higher than 10% difference in the Ki-67 proliferation index. When only the annotation with the highest positivity index from each patient was considered, five out of 34 cases showed over 10% difference. Two cases had higher than 15% difference. The four cases with the largest Ki-67 differences between the two methods are displayed in Figure 3.

**Relationship between metastasis-free and overall survival and Ki-67 expression.** The automated method associated a high Ki-67 status with worse overall survival [hazard ratio (HR)=1.047, p=0.024; 95% CI=1.006-1.090], but not the
visual method (HR=1.063, \(p=0.053\); 95% CI=0.999-1.131). Both automated (HR 1.030, \(p=0.118\); 95% CI=0.992-1.070) and visual (HR=1.034, \(p=0.254\); 95% CI=0.976-1.096) methods showed increased hazard ratio for metastasis but failed to achieve statistical significance. Kaplan-Meier survival curves (overall survival and metastasis-free survival) for both visually and automatically assessed highest Ki-67 positivity indexes classified into tertiles are shown in Figure 4.

**Discussion**

The main findings of the present study were the high agreement between visual and automated assessment of Ki-67 in synovial sarcoma, and the significant association between automated Ki-67 scoring and overall survival. Several clinical and histopathological features, such as tumor size and location, histologic grade, margin status, age, sex, and bone and vascular invasion have been studied to evaluate their potential prognostic value in synovial sarcoma. Tumor size over 5 cm has been consistently shown to be associated with shorter survival (14, 15). Also, other tumor related features such as undifferentiated morphology, high histologic grade, high mitotic rate and necrosis have been associated with poor prognosis (16, 17). Besides mitotic counting, measurement of the Ki-67 antigen by immunohistochemistry is the most widely used method for assessing the proliferative activity of tumors. Its expression can be visualized by immunostaining for Ki-67 with the MIB-1 antibody allowing the use of formalin-fixed paraffin-embedded (FFPE) material (18). Several previous studies have suggested the use of Ki-67 as a potential prognostic marker in soft-tissue sarcomas (19-22), including synovial sarcoma (20-22). The role of Ki-67 as a prognostic factor in STS is still conflicting; some studies have reported poor correlation between disease-free survival and Ki-67 status (23) whereas others have found strong correlation (24, 25). Skytting et al. (24) showed that MIB-1 index and tumor size are strongly associated to metastasis-free survival in synovial sarcoma. Hasegawa et al. (26) demonstrated that a grading system using Ki-67 index instead of counting mitoses was significantly associated with the prognosis of the patients with synovial sarcoma. The Ki-67 grading system performed also better than the FNCLCC grading system in terms of validity and reproducibility using mitotic score (27). Nevertheless, clinical adoption of a new histopathological parameter is challenged by the absence of harmonized methodology. Ki-67 scoring by visual assessment has been shown to suffer from considerable inter-observer and intra-observer variability in terms of positivity threshold interpretation, field selection and other factors (28, 29). International Ki-67 in Breast Cancer Working Group formulated evidence-based recommendations on both preanalytical and analytical assessment of Ki67 in 2011 (30). Consensus also included recommendations on interpretation and scoring of Ki67.

Digital image analysis techniques introduced a few decades ago have shown potential for automated assessment and possibly increased precision, but there are still limited data available on automated Ki-67 assessment in soft-tissue sarcoma. Inspired by the consensus recommendations, Acs et al. (9) used 30 ER+ breast cancer cases from the Working Group initiatives for standardization in their inter-platform...
and inter-operator study on Ki67 reproducibility using digital image analysis. For outcome cohort, samples from 149 breast cancer cases were used to build a TMA. The group demonstrated intraclass classification values of 0.964-0.970 between the three tested digital image analysis platforms and the reference standard Ki67 values of Spectrum Webscope. One of the tested platforms was QuantCenter also used in the present study. Inter-operator reproducibility also was high, and all tested platforms performed similarly when assessing breast cancer prognosis by using Ki67 (9).

The automated approach has some advantages over visual scoring: it counts cells faster and may increase precision (31). Although DIA may provide more detailed information in terms of cell calculation and reduce intra-observer and inter-observer variability, considerable care is required to establish reliable clinical measurements. For instance, tumor heterogeneity, tissue and staining artifacts, out-of-focus scanning, selection of representative area for counting, especially in tumors showing Ki-67 positive “hot spots” and adjusting the algorithm parameters are issues that still often require a human observer. In clinical practice, whole slide images are used instead of highly selected representative TMAs used in the present study. Therefore, deep learning algorithms are required to define the area of interest and to label all cells. In a comprehensive automated quantitative analysis on 1,017 whole slide images representing invasive breast ductal carcinomas, the accuracy, sensitivity, and specificity of identifying tumour regions were 89.44, 85.05 and 95.23%, respectively (10). Automatic calculation of Ki67 index showed accuracy of 90.2%, and when tested against the participating doctors the software showed an accuracy rate of 99.4% (10). One possible explanation for contradicting results on the correlation between Ki67 index and survival rates is a non-uniform selection of hot spot areas. Convolutional neural networks were exploited when testing three different methods of detecting Ki67 hot spot areas in whole slide images in invasive breast carcinoma (11). The method yielding best results with an overall accuracy of 95% used mutual information acquired from color deconvolution of both H&E and Ki67 stains.

In this study, images were representative with little inflammatory infiltrate and tumor heterogeneity (particularly monophasic subtype) and samples were selected from mitotically active areas. In some cases, however, it seemed

Figure 4. Survival curves by Ki-67 tertiles. Survival curves for the automated assessment method of the three Ki-67 subgroups: low<3.2%, moderate 3.2-14%, high>14.7% and for the visual assessment method: low<2.5%, moderate 2.5-11.3%, high>11.3%. The high tertile is illustrated with a thick line, the moderate with a thin line, and the low with a stippled line.
that the hypercellular, fascicular architecture with overlapping cells typical for synovial sarcoma was challenging for the algorithm in terms of tumor cell segmentation, in particular if there was even a slight blurriness in scanning. Therefore, the observed discrepancies between automated and manual tumor cell counts may be due to tumor cell under-segmentation as in many cases more than one cell were marked as a single object by the algorithm. The spindle-cell morphology typically found in synovial sarcoma might also have caused some of the disparities between the number of negative nuclei counted by the automated and manual method. As the size of the counted nuclei as well as the sensitivity of the algorithm to detect faintly stained nuclei can be adjusted by the software user, it is crucial to find appropriate settings. To overcome challenges concerning ki-67 nuclear detection and segmentation an automated multistep approach was utilized by a group in Milan, Italy (8). It is possible that weak counterstaining may lead to an overestimation of the Ki-67 positive cells (30), which may have been one reason for the somewhat higher scores obtained in the present study by the automated method. Therefore, it is possible that a stronger counterstain might have enabled the algorithm to better recognize negative nuclei. Furthermore, staining shows heterogeneity between and within tumour cells in terms of intensity and distribution. To study this heterogeneity Ewing sarcoma tumour cells were stained for DAPI and CD99 and several protein biomarkers (including Ki67) (7). Image feature distributions of different stains were then generated of each cell using automated image analysis, and the high dimensional data was integrated with clinical data utilizing random survival forest analysis to yield prognostic classifiers. This approach enabled classification of a subgroup of Ki67 positive tumour cells with low nuclear/cytoplasmic ratio of CD99 showing strong prognostic value.

Presently Ki67 index is not routinely used in soft-tissue sarcoma prognostic models. The choice of an optimal cut-off value for practical clinical use is also to be defined. Our Kaplan-Meier curves indicate that cases with Ki-67 scores in the lowest tertile had an excellent prognosis especially with moderate and high scores. This may naturally be due to chance in this small study, but, like the determination of optimal cut-off, must be investigated in a larger patient cohort. Despite these potential sources of error, we found that the mean Ki-67 positivity index measured by the automated and visual methods were similar, and highly correlated. The DIA software that we used, estimated slightly higher Ki-67 scores than manual counting. Our results are in agreement with Hasegawa et al. (26) who showed that image analysis-based MIB-1 LI was highly correlated with microscopic observation-based MIB-1 LI (r=0.87) in a series of 146 patients with soft-tissue sarcoma, including 44 synovial sarcomas. The significant association between automated Ki-67-scoring and prognosis, slightly better than between visual scoring and prognosis despite the high number of nuclei counted, indicate that the advantages of automated scoring may outweigh its drawbacks. Strengths of the present study include a homogenously treated patient material with long follow-up. A potential limitation to our study is the small number of patients. The use of TMA instead of whole sections might also be considered a limitation. The selection of a representative tumor area may have affected the results, as there is inevitably some heterogeneity in synovial sarcomas. Intra-tumor Ki-67 expression can vary significantly, and in a study concerning biphasic synovial sarcoma, Lopes et al. (32) observed higher expression of Ki-67 in the solid-glandular component than in the spindle cell component. Accordingly, we observed variability in our material as well, as there were annotations with positivity index ranging from 2-29% within the same specimen. When TMA technology was introduced in 1998 (33), it presented several advantages over whole-tissue sections, i.e. conserving reagents, saving time and preserving archived tissue, but it has been criticized for diminished reproducibility as the minute sample of the tissue may not be representative of the entire specimen. Nevertheless, several studies comparing TMA sections with whole-tissue sections have generally shown good agreement between the methods, also including soft tissue sarcomas (21, 34).

The study shows a good agreement between automated and visual counting of Ki-67 positivity index in synovial sarcoma. Ki-67 scores determined by the automated method were significantly associated with overall survival. Further validation in a larger dataset is required.

Conflicts of Interest

The Authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors’ Contributions

RL, TB, CB, ET and MS conceived and designed the study. MS, JR contributed to the analysis and interpretation of the data and validation in a larger dataset is required.

Acknowledgements

The study was supported by the Competitive Research Funding of Helsinki University Hospital and the Finnish Cancer Society.

References


Received July 6, 2021
Revised November 14, 2021
Accepted November 15, 2021