Ki67 Quantitative Interpretation: Insights using Image Analysis

Zoya Volynskaya[1,2], Ozgur Mete[1,2], Sara Pakbaz[1,2], Doaa Al-Ghamdi[3], Sylvia L. Asa[1,2]

[1]Department of Pathology, Laboratory Medicine Program, University Health Network, [2]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada, [3]Department of Pathology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Background: Proliferation markers, especially Ki67, are increasingly important in diagnosis and prognosis. The best method for calculating Ki67 is still the subject of debate. **Materials and Methods:** We evaluated an image analysis tool for quantitative interpretation of Ki67 in neuroendocrine tumors and compared it to manual counts. We expanded a primary digital pathology platform to include the Leica Biosystems image analysis nuclear algorithm. Slides were digitized using a Leica Aperio AT2 Scanner and accessed through the Cerner CoPath LIS interfaced with Aperio eSlideManager through Aperio ImageScope. Selected regions of interest (ROIs) were manually defined and annotated to include tumor cells only; they were then analyzed with the algorithm and by four pathologists counting on printed images. After validation, the algorithm was used to examine the impact of the size and number of areas selected as ROIs. **Results:** The algorithm provided reproducible results that were obtained within seconds, compared to up to 55 min of manual counting that varied between users. Benefits of image analysis identified by users included accuracy, time savings, and ease of viewing. Access to the algorithm allowed rapid comparisons of Ki67 counts in ROIs that varied in numbers of cells and selection of fields, the outputs demonstrated that the results vary around defined cutoffs that provide tumor grade depending on the number of cells and ROIs counted. **Conclusions:** Digital image analysis provides accurate and reproducible quantitative data faster than manual counts. However, access to this tool allows multiple analyses of a single sample to use variable numbers of cells and selection of variable ROIs that can alter the result in clinically significant ways. This study highlights the potential risk of hard cutoffs of continuous variables and indicates that standardization of number of cells and number of regions selected for analysis should be incorporated into guidelines for Ki67 calculations.

Keywords: Algorithm, continuous variables, digital pathology, Ki67, quantitative analysis, whole-slide imaging

Introduction

Subclassification of the diagnosis of several types of tumors, including breast cancers, brain tumors, adrenal cortical carcinomas, thyroid cancers, and neuroendocrine neoplasms, has been based on the quantitation of proliferation.[1-8] The classification and risk stratification of these diagnostic entities include mitotic counts and the assessment of a Ki67 labeling index.[9] Ki67 was initially identified as an antigen associated with mitosis in mammalian cells by investigators in Kiel (hence the Ki in the name).[10] The use of this biomarker has become the subject of intense controversy.[6] The labeling index of this antigen has been counted by eyeballing slides, by manual counts of printed images photographed at the microscope, and by automated image analysis algorithms. Because the reproducibility of Ki67 positive cell counts is poor, particularly when eyeballing[7,11-13] careful manual counts of printed images or automated image analysis have been recommended to improve the accuracy of this biomarker.[12-14] In addition, staining results are subjected to interlaboratory variation that is dependent on both tissue fixation and staining technology.[8,15]

In an effort to ensure accurate and reproducible Ki67 labeling indices, we implemented an image analysis tool in the Department of Pathology at the University Health Network, Toronto. The validation was undertaken by the endocrine pathologists who assess a large number of cases, for which the Ki67 labeling index is used to grade neuroendocrine neoplasms.[3,14] During the course of validation, we compared this tool to the previous method of calculating the Ki67 labeling index, manual counts of printed images of the region

Access this article online	
Quick Response Code:	**Website:** www.jpathinformatics.org
	DOI: 10.4103/jpi.jpi_76_18

of interest (ROI). We report here the results of this validation in terms of accuracy, time, and reproducibility. More importantly, during this validation, it became apparent that different types of specimens, specifically biopsies or resections, alter the availability of tissue for analysis, and some biopsies did not yield the recommended number of cells; the availability of this tool allowed comparisons of different ROIs based on the number of cells and number of regions selected for analysis.

MATERIALS AND METHODS

Materials

According to guidelines,[16] following primary use-case validation of digital pathology using at least 60 cases, each additional use-case validation requires 20 additional cases. For this study, we collected 20 consecutive cases of neuroendocrine neoplasms; these tumors had Ki67 labeling indices reflective of the wide range of these tumors, from very low (approximately 0.1%) to high (approximately 75%). These included primary neuroendocrine tumors of stomach, small bowel, appendix, pancreas, lung and ovary, liver metastases from lung and small bowel neuroendocrine tumors, and paraganglioma. Sections of 5-μm thickness were stained on the Roche Ventana Benchmark using the MIB1 antibody (Dako, Santa Clara, CA, USA). Slides were scanned with a Leica Aperio AT2 Scanner (Leica Biosystems, Vista, CA, USA) and accessed through the CoPathPlus laboratory information system (Cerner, Kansas City, MO, USA) interfaced with Aperio eSlideManager through Aperio ImageScope (Leica Biosystems) as previously described.[17] The pretuned nuclear algorithm (Leica Biosystems) was used for automated analysis of slides stained for Ki67.

Validation of image analysis algorithm

Since the program used does not have the ability to identify the regions of highest labeling, also known as "hotspots," we identified ROIs on the digital slides by visually selecting the area of highest labeling. We outlined the ROI using a frame and then annotated the area within the frame using the ImageScope software to manually outline stroma for exclusion in the ImageScope analysis [Figure 1]. The selected and annotated regions were photographed, printed, and distributed to four pathologists (OM, SP, DAG, and SLA) who performed manual counts as per their usual practice following the WHO recommendation that "manual counting using printed images is advocated."[14] The outlined ROIs in the whole slide image were subjected to image analysis using the image analysis nuclear algorithm for determination of the Ki67 labeling index on three occasions. Each analysis was timed from the onset of analysis to completion. This timing did not include the time required for ROI selection and annotation, as the annotations were made in advance and then the ROI was printed ×4 for manual counting by each individual who was blinded to the results of the other users and the image analysis algorithm.

The results of the analyses by each of the four pathologists and the algorithm were compiled [Table 1] and compared [Figure 2].
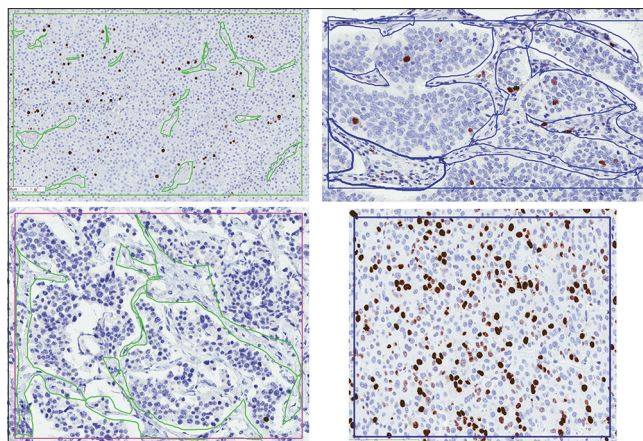

Figure 1: Sample photographs of annotated figures used for Ki67 quantitation by manual counts and automated image analysis. These images illustrate examples of how annotations were applied for quantitation of Ki67 labeling index for validation of an automated algorithm. Sample slides were annotated with a square to identify the region of interest, then annotations were made to exclude the stroma. The resulting images were printed for manual counting and subjected to the automated algorithm for analysis

Reproducibility of the algorithm based on user determinations

Since some of the specimens were resection specimens with large tissue pieces, while others were biopsies with fewer cells, and some of the biopsies were intact cores, while others were multiple small fragments, we recognized that the selection of a "hot spot" varied from specimen to specimen. In resection specimens, the initial approach was to identify the single area of highest labeling and identify an area that had at least 1000 cells; in some cases, more than 1000 cells were counted. In some biopsies, it was difficult to obtain 1000 cells, and in fragmented biopsies, to achieve a total cell count of more than 500 cells, it was frequently necessary to identify multiple small "hotspots."

To determine the impact of selection of the ROI on the outcome and classification of the tumor, we performed the analysis using multiple approaches. We carried out the automated analyses on a single tumor using different numbers of total cells and multiple small versus single large areas of hot spots.

RESULTS

Validation of image analysis algorithm

The results obtained using the automated image analysis algorithm compared with manual counts of the same annotated areas by multiple observers are shown in Table 1 and illustrated in Figure 2.

Analysis of annotated areas by the algorithm resulted in identical results when repeated multiple times. Overall, the automated analysis correlated with the pathologists' results in the majority of cases. However, the manual counts varied from pathologist to pathologist [Figure 2] depending on the interpretation of the selected area. Specifically, there were two areas of difference. First, while the algorithm identifies any

Table 1: Comparisons of results of Ki67 in 20 cases

Cases	Nuclear algorithm		Observer 1			Observer 2			Observer 3			Observer 4		
	Percentage	Total # cell	Percentage	Total # cell	Time, min	Percentage	Total # cell	Time, min	Percentage	Total # cell	Time, min	Percentage	Total # cell	Time, min
Case 1	2.119	1463	1.84	1412	10	1.57	1400	30	1.99	951	9	1.76	1303	Time was not recorded for individual cases. Estimated time was between 15 and 30 min for each case
Case 2	0.107	931	0.12	853	8	0.17	574	13	0.16	618	7	0.12	799	
Case 3	3.425	1664	3.84	1951	20	2.26	1944	50	6.53	826	10	2.91	1747	
Case 4	4.145	3643	5.46	2380	22	3.85	2392	48	7.08	1581	20	4.49	2334	
Case 5	6.586	911	6.98	1147	23	5.44	1047	15	12.15	576	9	3.18	2008	
Case 6	1.606	2366	2.09	956	23	1.67	836	16	2.98	569	8	1.92	727	
Case 7	1.497	1670	1.25	1600	22	1.44	1320	23	2.13	935	7	1.54	1422	
Case 8	3.091	2556	3.89	2186	32	2.56	2104	45	7.68	924	15	3.75	1810	
Case 9	2.870	1289	3.06	1177	25	2.32	1290	20	4.06	689	10	3.11	1123	
Case 10	9.287	1346	11.31	1494	18	7.36	1522	40	13.23	922	13	9.66	1314	
Case 11	2.027	1529	1.92	3122	28	1.16	2681	55	2.27	1540	21	1.58	2423	
Case 12	5.849	1915	7.47	2125	22	4.88	1945	30	13.22	809	13	7.61	1510	
Case 13	29.651	1059	31.02	1054	25	27.31	908	18	38.12	758	14	32.4	876	
Case 14	1.971	558	2.11	427	9	2.76	326	6	2.76	289	3	2.44	368	
Case 15	47.259	821	48.85	782	18	36.11	720	15	56.35	456	9	45.92	699	
Case 16	5.151	1553	5.77	1525	20	3.97	1510	28	8.71	769	11	5.43	1269	
Case 17	10.779	1375	9.55	1476	27	7.47	1405	22	15.29	817	11	9.35	1283	
Case 18	1.692	2601	1.63	1778	25	1.32	1216	25	3.61	831	11	1.9	1315	
Case 19	19.350	571	15.29	713	12	10.41	826	13	22.52	444	10	14.48	773	
Case 20	73.374	661	81.24	485	10	62.30	573	13	86.64	614	9	81.74	586	

staining as positive, there was discordance among pathologists regarding inclusion of very weak signals. Second, some pathologists counted all nuclei, while others omitted nuclei within the analysis area that could possibly be interpreted as stroma. These differences are known to contribute to interobserver variability in tumor grading.

Time savings by image analysis

The automated algorithm provided results within few seconds compared to up to 55 min per analysis when performed manually [Table 1]. This timing did not include the time required for annotation of the initial image; it only included the time for actual counting on printed images.

Reproducibility of the algorithm based on user determinations

We repeated the analysis of a given tumor using larger or smaller frames to include more or fewer cells in the same region that had been identified as the "hotspot." We identified a consistent variation of the Ki67 labeling index based on cell number; the more cells counted, the lower the Ki67 value obtained, with the highest variation at the low end of the spectrum [Figure 3]. This impacted the cutoff points that have been defined for the distinction of Grade 1 from Grade 2 neuroendocrine tumors[3,14] such that counting 1000 cells could result in a tumor being classified as moderate grade (G2), whereas counting 2000 cells or more resulted in the same tumor being classified as low grade (G1). The same occurred for tumors close to the 20% cutoff for intermediate- versus high-grade (G3) classification.

Since biopsies can be fragmented and yield multiple small pieces of tissue for analysis, we then examined the impact of selecting multiple small regions compared to a single large region to obtain the required number of cells [Figure 4]. It was evident that selection of multiple regions of intense labeling resulted in a higher value than selection of a single region of the same number of cells even if overall that represented the most intense hot spot.

CONCLUSIONS

This study was performed as part of the validation of new technology in the laboratory. Our results confirm that there is a significant benefit of automated image analysis as part of daily pathologists' workflow, both in the consistency of the automated results and in the time savings for pathologists. Our study did not include the time required to identify hotspots or to annotate images for counting, we assume that the time required for such annotation would not be significantly different using printed images or the ROI on a computer screen since the work to do this is mainly a factor of pathologist recognition and labeling of stromal elements for exclusion. As new algorithms are developed that can recognize hot spots and perform automated segmentation to exclude stromal elements, the time required for these activities will be reduced.

The ability to perform fast and reliable image analysis for quantification of morphologic features allowed us the opportunity to pursue a more in-depth analysis of the impact of tissue annotation for the analysis of the Ki67 labeling index.
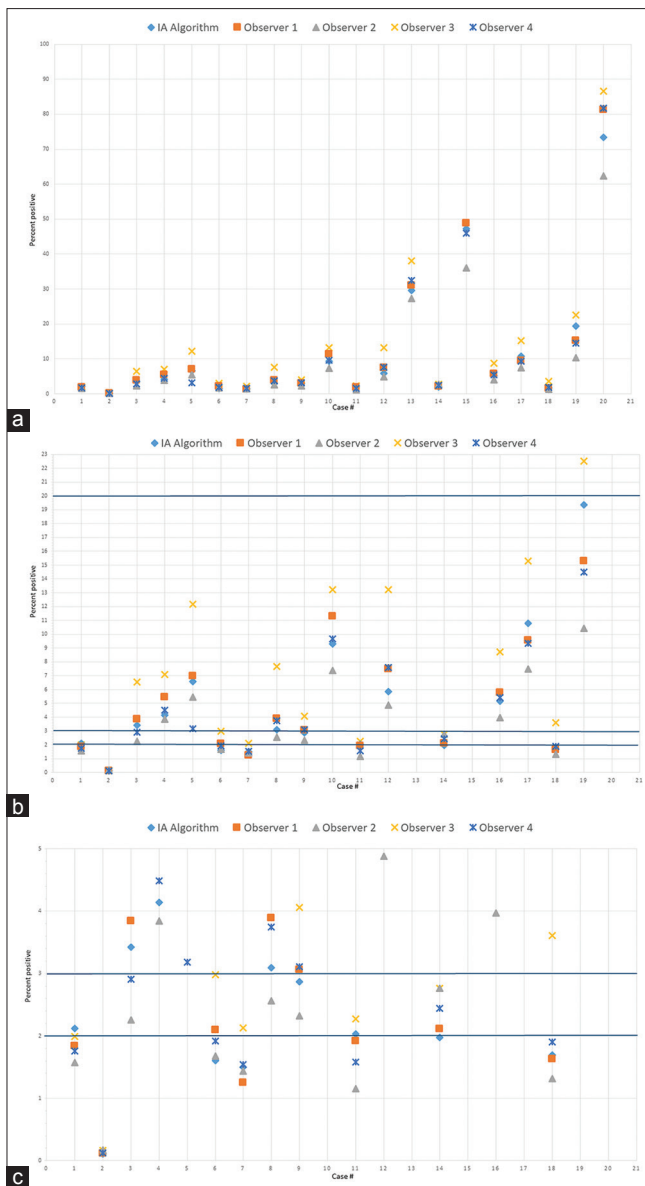
Figure 2: Results of manual and automated counts of Ki67 labeling index. (a) Includes the entire scale from 0% to 100%. (b) Is an expanded view of the cases close to the 20% cutoff for G2 versus G3 neuroendocrine tumors. (c) Is expanded to show the variability at the previous 2% and revised 3% cutoff to separate G1 from G2 neuroendocrine tumors

We have identified significant interobserver variation due to pathologist interpretation. Despite the instructions to use the prepared annotations, one pathologist had consistently higher Ki67 results when using manual counts of printed images. This was attributed to the fact that this pathologist excluded any cell that was perceived as stroma or blood vessel even if it was included in the countable region of the ROI. Therefore, the results would have differed whether using manual counts or automated image analysis. Some differences may possibly be related to the inclusion of very weak signals by some and not all observers, the algorithm is set to include even weak and/or focal staining as positive.

This study analyzed hotspots that were identified by a pathologist. As image analysis tools become more sophisticated,

Figure 3: Results of automated counts using different numbers of cells. In the example shown in (a), counting a field that included 1906 cells provided a Ki67 labeling index of 2.99%; in contrast, counting 1455 cells in the same area using a smaller region of interest resulted in a Ki67 labeling index of 3.09% for this tumor. Using the current WHO classification, the difference makes this either a G1 or a G2 tumor. In (b), counting 1421 cells provide a Ki67 of 19.49% and classification as a G2 tumor, whereas counting 992 cells results in a Ki67 of 21.27% and classification as a G3 tumor

automated tumor/stroma segmentation will become a common standard. However, the insights that we obtained in our study will need to be considered when developing those segmentation algorithms, to establish how to identify a hotspot based on the total number of cells available or required, whether a single area only should be analyzed, and if so, how such an algorithm can be applied to small biopsies.

While variation in Ki67 labeling results has been attributed to the known heterogeneity of different areas within neuroendocrine tumors,[18] we have also confirmed significant variation of the Ki67 labeling index based on the size and number of ROIs. Our data confirmed the obvious result that counting more cells skews the result for cases with relatively low Ki67 labeling. The same is true when selecting multiple very small areas compared with a large tissue region, even when the total number of cells is the same. This concentration effect has a significant impact when considering that much of current practice rests on results obtained from small biopsies that frequently have either too few cells for a complete analysis or may be fragmented, yielding multiple small regions. The ability to push a tumor from a G1 to G2 or G2 to G3 classification can be as simple as reducing the number of cells counted from 1500 to 1000 or selecting multiple small hot spots rather than a single larger area of the same tumor. The literature has not dealt with this issue rigorously; the WHO has recommended that "the Ki67 proliferation index is based on the evaluation of equal and >500 cells in areas of higher nuclear labeling (so-called
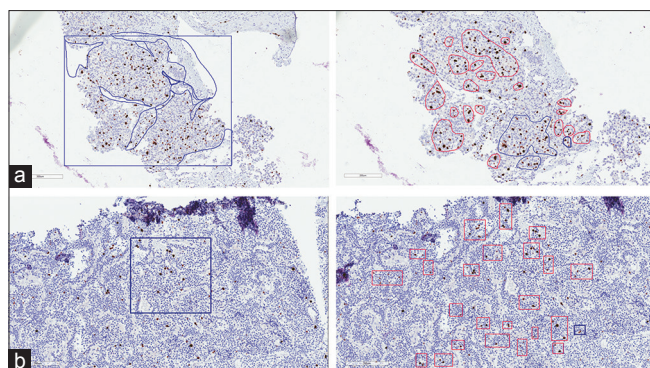
Figure 4: Results of automated counts using multiple regions of interest versus a single region of interest. The figures on the left show the annotation of a large area, whereas those on the right show multiple small areas selected for analysis. These different approaches to annotation of the small biopsy in (a) yielded a Ki67 labeling index of 13.2829% in 1453 cells (left) versus 21.825% in 996 cells (right). In the larger sample shown in (b), the results were 2.185% in 1144 cells (left) versus 7.285% in 1057 cells (right)

hotspots)."[14] The implication of this is that more is better, and some pathologists try to count as many cells as possible; with automated tools, it is easier to count more cells or even the entire slide, yet our data show that this alters the result in a way that can be significant to grading of a neuroendocrine tumor. Careful studies based on rigorous and consistent protocols are needed to prevent concentration or dilution effects and to determine the correct mechanism for counting that is clinically relevant. The application of this analysis in biopsies complicates the matter since often biopsies do not contain sufficient numbers of cells to achieve recommended counts. While some would argue that results close to a cutoff should be rounded up to the nearest whole integer,[14] there are no guidelines on this issue. There is a need for a rigorous study using image analysis tools and standardized numbers of cells to determine the diagnostic cutoffs that are clinically significant or to update the approach to this continuous variable. Indeed, it may be that the Ki67 labeling index should not have set cutoffs that can be manipulated as showed in this study. The impact of this on grading of neuroendocrine tumors and other tumor types will be significant.

In conclusion, we report that the application of automated image analysis for the enumeration of a Ki67 labeling index provides a fast and accurate tool for this methodology. However, we provide examples of variations that result from the size and number of selected fields to determine the ROI for analysis. These results highlight the importance of developing a standardized approach to quantitation in anatomical pathology and raise concerns about the rigid cutoffs for tumor grading that have been promoted based on nonstandard studies.

Financial support and sponsorship

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Louis DN, Edgerton S, Thor AD, Hedley-Whyte ET. Proliferating cell nuclear antigen and Ki-67 immunohistochemistry in brain tumors: A comparative study. Acta Neuropathol 1991;81:675-9.
2. Tallini G, Garcia-Rostan G, Herrero A, Zelterman D, Viale G, Bosari S, *et al.* Downregulation of p27KIP1 and Ki67/Mib1 labeling index support the classification of thyroid carcinoma into prognostically relevant categories. Am J Surg Pathol 1999;23:678-85.
3. DeLellis RA, Lloyd RV, Heitz PU, Eng C. Pathology and Genetics of Tumours of Endocrine Organs. Lyons, France: IARC Press; 2004.
4. McCall CM, Shi C, Cornish TC, Klimstra DS, Tang LH, Basturk O, *et al.* Grading of well-differentiated pancreatic neuroendocrine tumors is improved by the inclusion of both Ki67 proliferative index and mitotic rate. Am J Surg Pathol 2013;37:1671-7.
5. Singh S, Hallet J, Rowsell C, Law CH. Variability of ki67 labeling index in multiple neuroendocrine tumors specimens over the course of the disease. Eur J Surg Oncol 2014;40:1517-22.
6. Polley MY, Leung SC, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JM, *et al.* An international study to increase concordance in Ki67 scoring. Mod Pathol 2015;28:778-86.
7. Papathomas TG, Pucci E, Giordano TJ, Lu H, Duregon E, Volante M, *et al.* An international Ki67 reproducibility study in adrenal cortical carcinoma. Am J Surg Pathol 2016;40:569-76.
8. Focke CM, Bürger H, van Diest PJ, Finsterbusch K, Gläser D, Korsching E, *et al.* Interlaboratory variability of Ki67 staining in breast cancer. Eur J Cancer 2017;84:219-27.
9. Rudolph P, Peters J, Lorenz D, Schmidt D, Parwaresch R. Correlation between mitotic and Ki-67 labeling indices in paraffin-embedded carcinoma specimens. Hum Pathol 1998;29:1216-22.
10. 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;31:13-20.
11. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS. Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: A comparison of digital image analysis with manual methods. Am J Surg Pathol 2012;36:1761-70.
12. Young HT, Carr NJ, Green B, Tilley C, Bhargava V, Pearce N, *et al.* Accuracy of visual assessments of proliferation indices in gastroenteropancreatic neuroendocrine tumours. J Clin Pathol 2013;66:700-4.
13. Reid MD, Bagci P, Ohike N, Saka B, Erbarut Seven I, Dursun N, *et al.* Calculation of the Ki67 index in pancreatic neuroendocrine tumors: A comparative analysis of four counting methodologies. Mod Pathol 2015;28:686-94.
14. Lloyd RV, Osamura RY, Kloppel G, Rosai J. WHO Classification of Tumours of Endocrine Organs. 4th ed. Lyon: IARC; 2017.
15. Blank A, Wehweck L, Marinoni I, Boos LA, Bergmann F, Schmitt AM, *et al.* Interlaboratory variability of MIB1 staining in well-differentiated pancreatic neuroendocrine tumors. Virchows Arch 2015;467:543-50.
16. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, *et al.* Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of American Pathologists Pathology and Laboratory Quality center. Arch Pathol Lab Med 2013;137:1710-22.
17. Volynskaya Z, Chow H, Evans A, Wolff A, Lagmay-Traya C, Asa SL, *et al.* Integrated pathology informatics enables high-quality personalized and precision medicine: Digital pathology and beyond. Arch Pathol Lab Med 2018;142:369-82.
18. Yang Z, Tang LH, Klimstra DS. Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: Implications for prognostic stratification. Am J Surg Pathol 2011;35:853-60.