Short Communication

Optimal settings and clinical validation for automated Ki67 calculation in neuroendocrine tumors with open source informatics (QuPath)

Rima Pai [a], Susan Karki [b], Rakhee Agarwal [c], Steven Sieber [b], Samuel Barasch [b,*]

[a] Nuvance Health Department of Research and Innovation, 14 Research Dr., Bethel, CT 06801, USA
[b] Nuvance Health, Danbury Hospital, Department of Pathology and Laboratory Medicine, Tower building, 2nd floor, Lab, 24 Hospital Ave., Danbury, CT 06810, USA
[c] North America Clinical Functional Service Provider, IQVIA, Parsippany, NJ 07054, USA

Introduction

Open source informatics was sought to reduce intra/interobserver variability in the calculation of Ki67 percentage in neuroendocrine tumors. After choosing the "Positive Cell Detection" function in QuPath,[1] a study was conducted to assess optimal settings, frequency of setting variation, and data was collected for validation of the platform as a potential clinical decision-support tool.

Technical background

Image analysis is a growing field of informatics which provides clinical decision-support to physicians and others working with diagnostic imagery. A solution which was reliable and low cost was sought to begin studying workflow related to image analysis for Ki67 percentage in neuroendocrine tumors, data which is used to grade these tumors and predict recurrence or metastasis in patients. Image analysis on computers works through a mathematical separation of colors in an image file. The colors in the computer image file are determined by assigning a numerical value to each pixel which represents the balance between the colors red, green, and blue at that point (pixel) in the image. Because a definitive number is assigned to each pixel in each image, there is potential for a computer to parse image colors by an automated method (an algorithm which defines a function). Due to histologic variability (slide staining, tissue thickness, etc.), it is necessary, with some solutions, to recalibrate the vectors which separate the red, green, and blue colors (i.e. pixel depth numbers) for each image.

Because the images of the immunohistochemical stains can be numerically quantified, the automated process with established thresholds for positivity can be used to collect high quality, precise data by limiting inter- and intra-observer variability.[2,3] Prior studies have used image analysis to evaluate antibody clones used in immunohistochemistry.[4,5]

To keep costs low during study of this technology, an open source solution was sought. 'Open source' is a description of software in which the instructions given to the computer to complete a function (an algorithm written in code) are made freely available. However, open source products can frequently require more training or user expertise to implement or use effectively. Therefore, authors have sought to develop a group of Optimal Settings which would require minimal fine tuning of settings when different immunohistochemical stains from different tumors are analyzed. For example, our goal in identifying Optimal Settings was to develop a group of software settings which would give an accurate percentage positivity under the Positive Cell Detection function (QuPath) in most circumstances without adjusting the settings. These Optimal Settings would reduce the amount of expertise needed to use Positive Cell Detection function and ensure that most users, for most tumors, would achieve an accurate count on their first effort. In order to assess how much expertise would be required to use the open source informatics, optimal settings were developed and frequency of required deviation from optimal settings was tracked as a surrogate for level of technical acumen required to generate a Ki67 percentage with QuPath. Additionally, data was collected for validation of the method.

Prior validation of the package has been performed on tissue micro arrays with clinical outcome correlation, so this tool could be approached with confidence.[6,7]

Approach/procedure

This study was determined to be exempt by the institutional IRB. Procedure and optimal settings for using the 'Positive cell detection function' in QuPath are outlined in Table 1. Case selection for study and validation included neuroendocrine tumors of the gastrointestinal tract as well as metastatic neuroendocrine tumors presented at the Danbury Hospital Gastrointestinal tumor board over a period of 4 years which had a Ki67 immunohistochemical stain performed as part of the diagnostic pathology. Prior tumor board presentations were reviewed to identify cases for inclusion. Neuroendocrine tumors were selected independent of anatomic site. Images of Ki67 IHC slides were photographed at 10x with Olympus Infinitiy Capture 6.5.4. The images were stored as JPG files, and imported to QuPath 0.2.3 as "Heme/DAB brightfield" images. Images were then annotated (with appropriate polygon annotation avoiding background epithelium or inflammation) to approximate a 500 tumor cell count in Ki67 hotspots (i.e. areas of greatest immunolabeling of tumor cells for this proliferation

* Corresponding author at: Pathology and Laboratory Medicine, Tower building, 2nd floor, Lab, 24 Hospital Ave., Danbury, CT 06810, USA.
  E-mail addresses: Susan.karki@nuvancehealth.org (S. Karki), Steven.sieber@nuvancehealth.org (S. Sieber), Samuel.Barasch@nuvancehealth.org (S. Barasch).

Table 1

Procedure and optimal settings in QuPath for "Positive Cell Detection" function applied to Ki67 in neuroendocrine tumors.

Photomicrograph	10x, resolution 2080 × 1536 pixels, zoom 50%
Image file type	JPG
Image type (upon import to QuPath)	Heme/DAB brightfield
Pixel depth separation vectors	"Estimate Stain Vectors" function with "Auto" calibration (Fig. 1)
Heme threshold for counterstain	0.01 (NOTE: default is 0.1)
"thresholdCompartment"	Nucleus: DAB OD mean
"thresholdPositive1"	0.2 (NOTE: this is the default)

marker). Settings for photomicrograph magnification and resolution were held constant. Optimal settings were determined through "trial and error" to include the following: Photomicrograph at 10×, resolution of 2080 × 1536 pixels, with zoom of 50%, RGB pixel depth stain vectors were recalibrated with "Estimate Stain Vectors" before algorithm count using the default "auto" detection (Fig. 1). The heme threshold for counterstain "threshold" was changed from default setting (0.10) to 0.01; "thresholdCompartment" was set to be "Nucleus: DAB OD mean" with "singleThreshold" set to be "true"; "thresholdPositive1" was set to be "0.20" (the default).

Following the use of the Positive Cell Detection algorithm from QuPath, before and after images were reviewed in concert to assess the adequacy of
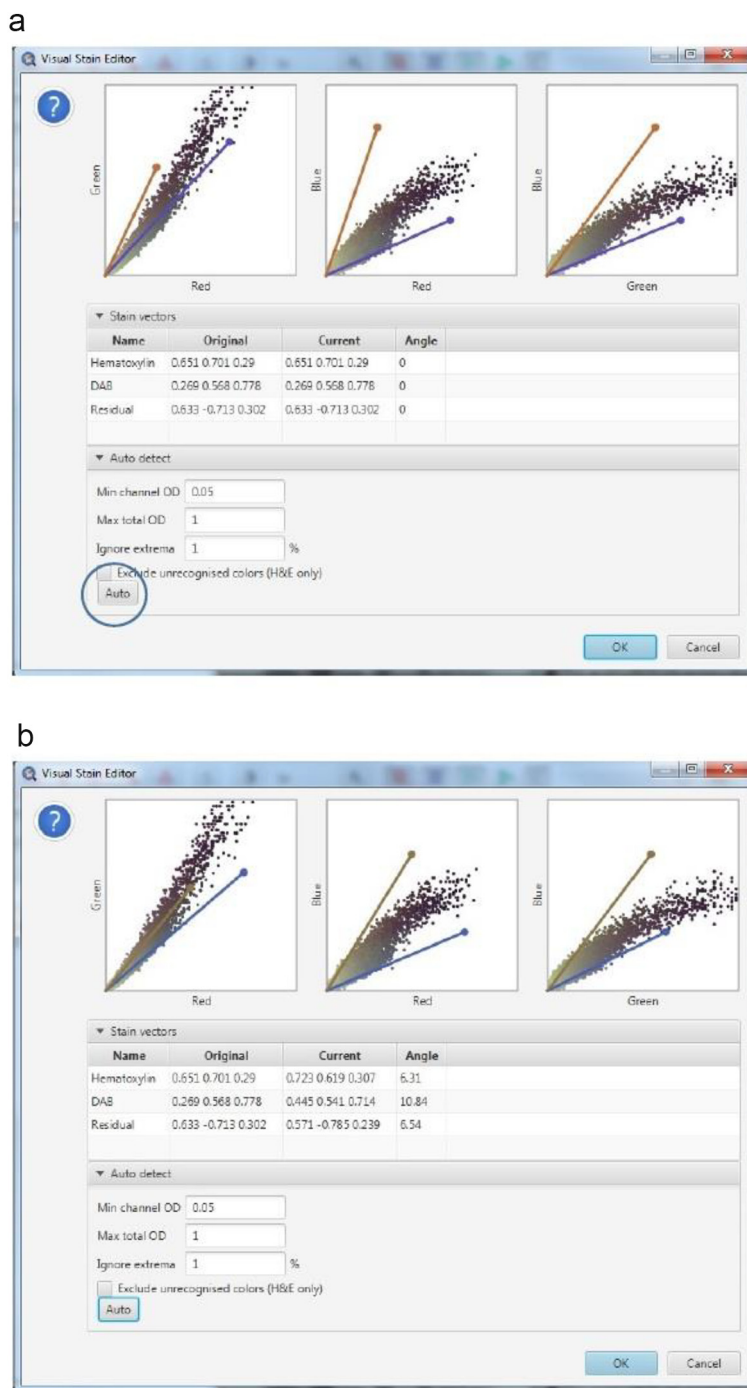
a

b

Fig. 1. Image auto-calibration via "Estimate Stain Vectors" function to correct for pixel depth variability between images.

thresholds (Fig. 2 – comparison of pre- and post-Positive cell Detection algorithm Ki67 images). In situations where adequate numbers of cells or positive cells were not counted by the automated system, thresholds were adjusted and repeat processing of image with new thresholds was performed. Frequency of changes required to achieve an adequate automated count of Ki67 positive immuno-labeled tumor cells was tracked (Table 2)

For validation, adequacy of positive threshold for each image was established through a second pathologist review of annotations with and w/out
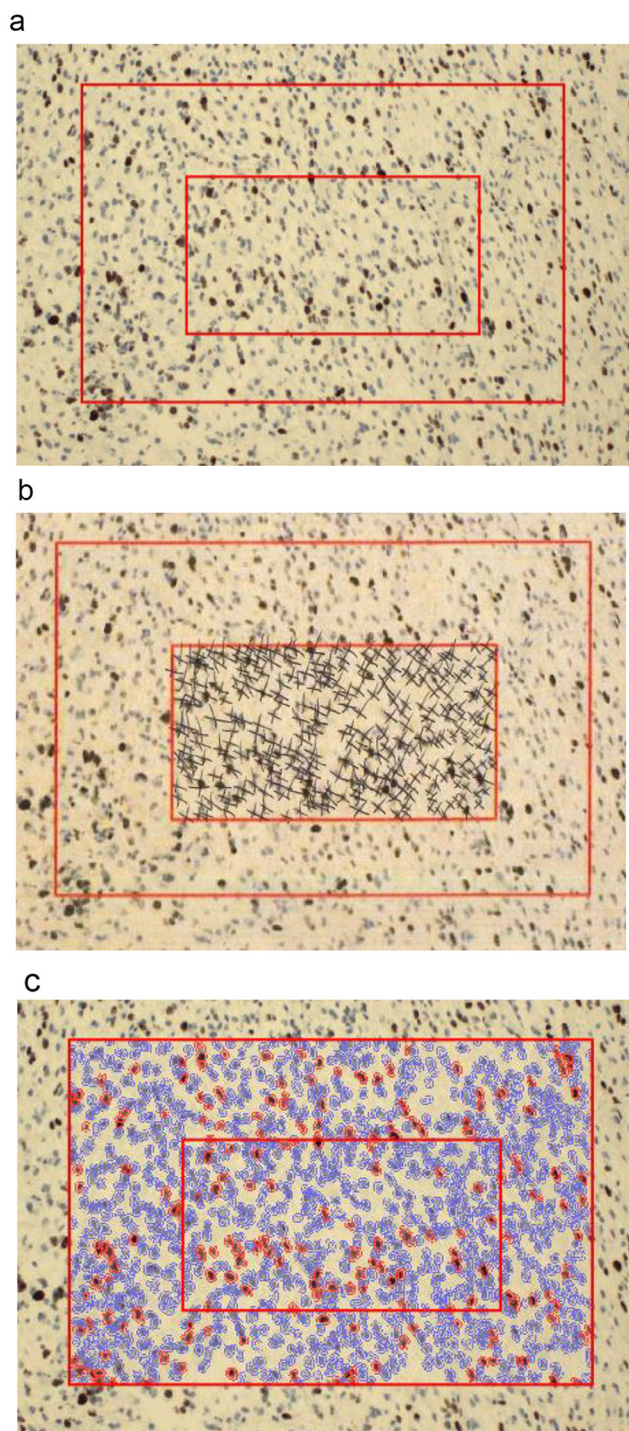

a

b

c

Fig. 2. Comparison of annotated image (2a) with manual cell count (2b) and automated cell count (2c). 2a) Annotated tumor image with Ki67 immunohistochemistry. 2b) Annotated tumor image with manual counting. 2c) Annotated tumor image with automated system (QuPath) counting.

highlights of positive and negative cells. Then, positive and negative cells in annotated areas were counted manually *blinded to the algorithm results* (percent positive calculated) and compared to the output from QuPath Positive Cell Detection (Fig. 2). MS Excel and R4.1.1 were used for graphs and statistics (paired T-test for cell count, ANOVA for grade difference – Table 3). For statistical analysis, tumors were grouped into grades 1, 2, and 3 by standard College of American Pathology Ki67 percentage positive for grades 1, 2, and 3 neuroendocrine tumors (i.e. Grade 1 NET is less than 3%, Grade 2 NET is 3–20%, and Grade 3 is >20% positive). Because Ki67 percentage is reported as whole numbers, rounding to the nearest whole number was performed before differences in grading between methods were assessed. Statistical analysis of differences in cell count were done without regard to grading or percentage positive.

Results

41 tumors were selected for analysis. Optimal settings for image capture were determined to be at 10x and described in more detail in PROCEDURE above and in Table 1. After determining optimal settings, 15% (6 of 41) of images required software settings changed to have adequate positive thresholds for automated counting and Ki67 proliferation index calculation in QuPath (1 case heme counterstain threshold changed to 0.1; 3 cases DAB stain compartment from mean to max; 5 cases had DAB positive threshold changes (0.23, 0.23, 0.30, 0.17, and 0.18). See Table 2.

Statistics for total cell count and Ki67 percentage differences are shown in Table 3. A statistical difference for cell count was present between methods. However, no significant statistical difference between tumor grading was identified between methods (alpha = 0.05).

A plot of difference between Ki67 percentage determined by automated and manual methods is seen in Fig. 3 (data skew allows shows both Ki67% by manual count and Ki67% by automated method and their comparison. Skew accommodates higher-grade tumors.) Four tumors Grade 1 manually were Grade 2 by QuPath; 1 tumor, which was Grade 2 manually, was Grade 1 by QuPath.

Conclusion

This is a moderately sized method validation for automated Ki67 percentage. Although the "site blind" nature of tumor type selection limits biological interpretation of these results, other authors have approached image analysis independent of tumor site.[8]

Prior studies have used vendor-dependent tools and this is the first study publishing optimal settings for this open source platform. The small number of times that the optimal settings (Table 1) need to be adjusted opens the door to this open source package being used in other studies and clinical applications with well controlled tissue types. The fact that changes to the optimal settings were uncommon are strong evidence that this is a reliable solution to automated Ki67 counting. With uncommon changes to settings, training staff (or residents) to use this package to collect high quality data on tumor staining while limiting intra- and inter-observer variability is attainable. Furthermore, this open source technology can be a

Table 2
Frequency of changes and types of changes required to optimal settings (Table 1) to achieve acceptable automated Ki67 enumeration.

Total number of tumor images counted: 41	Tumor images which needed settings changed: 6 (~15%)
Number of tumor images requiring change to Heme counterstain threshold	1 (Change from optimal setting of 0.01 to 0.1)
Number of tumor images requiring change to "thresholdCompartment"	3 (changed from "Nucleus: DAB OD mean" to "Nucleus: DAB OD max")
Number of tumor images requiring change to "thresholdPositive1"	5 (changed from 0.2 to 0.23, 0.23, 0.3, 0.17, and 0.18 respectively)

NOTE: Some tumor images required more than one adjustment to the optimal settings to achieve an acceptable Ki67%.

Table 3

Differences between total cell count and Ki67 % positive by manual and automated methods: Mean, median, and range.

	Range by manual method	Range by QuPath	Mean difference between methods	Median difference between methods	Minimum observed difference between methods	Maximum observed difference Between methods
Cell count (numbers of annotated cells)	242–774	270–865	120*	103*	0	389
Ki67% positive	0.2%–45%	0.2%–47%	1.02%**	0.55%**	0.4%	8.9%

*Difference statistically significant (p < 0.001); **No statistically significant difference in grading detected (p = 0.45).
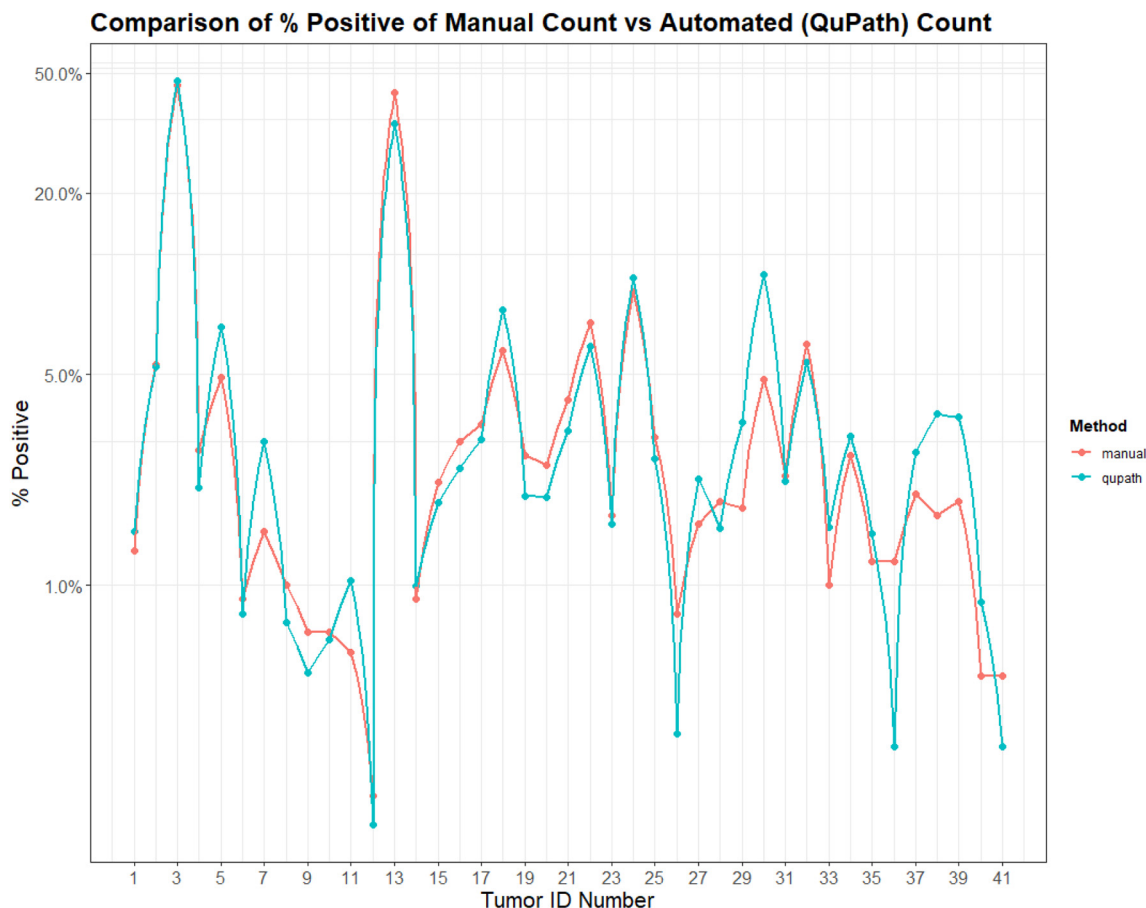

Fig. 3. Comparison of % Positive of Manual Count vs Automated (QuPath) Count.

useful bridge to adapting new software while seeking institutional and vender support.

With any image analysis package, image capture is a critical step (see the mantra "garbage in-garbage out"). Holding microscopic magnification to 10x was a critical step. Within the software, it is absolutely necessary to recalibrate each image with the "Estimate stain vectors" function. While understanding of the mathematical underpinnings of pixel representation of color within a JPG image file is not within the scope of knowledge for most pathologists or residents, almost everyone should be able to press the "Auto" button to calibrate the vectors (Fig. 1). After image calibration, more than 80% of the images in this set required no changes to the settings for optimal automated counting.

While no statistical difference between tumor grading was detected, the automated counting method seems to give a higher cell count. This is difficult to understand but may be due to variability in cell shape and size as represented in the standard histology slides. Despite variability in cell count between methods, the grade (more clinically relevant) was not found to be different between the automated and manual methods. Prior studies have found Ki67 to be diluted when more than 500 cells are counted; the same study expressed concern for potential over counting of Ki67 proliferative index by automated methods.[3,4]

Although all of the methodological differences in tumor grading were identified in Grade 1 and Grade 2 assessments of neuroendocrine tumors in this study, there were few Grade 3 neuroendocrine tumors available for study. Despite the fact that no difference between manual and automated methods was seen in our data, our findings might suggest that using an automated method as an adjunct in grading neuroendocrine tumors may be most helpful in lower grade tumors. Although studies with smaller data sets have found differences in grading at the Grade 1/Grade 2 decision point,[3,9] further studies which examine the utility of image analysis at this break point would be helpful.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7(1): 16878. https://doi.org/10.1038/s41598-017-17204-5.12 04.
2. Acs B, Pelekanou V, Bai Y, Martinez-Morilla S, Toki M, Leung S, et al. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. Lab Investig J Techn Methods Pathol 2019 Jan 2019;99(1). https://doi.org/10.1038/s41374-018-0123-7.
3. Luchini C, Pantanowitz L, Adsay V, Asa SL, Antonini P, Girolami I, et al. Ki-67 assessment of pancreatic neuroendocrine neoplasms: systematic review and meta-analysis of manual vs. digital pathology scoring. Modern Pathol Off J United States Can Acad Pathol Inc Jun 2022;35(6). https://doi.org/10.1038/s41379-022-01055-1.
4. Owens R, Gilmore E, Bingham V, Cardwell C, McBride H, McQuaid S, et al. Comparison of different anti-Ki67 antibody clones and hot-spot sizes for assessing proliferative index and grading in pancreatic neuroendocrine tumours using manual and image analysis - Owens - 2020 - Histopathology - Wiley Online Library. Histopathology 2020;77(4):646–658. https://doi.org/10.1111/his.14200.
5. Aung T, Acs B, Warrell J, Bai Y, Gaule P, Martinez-Morilla S, et al. A new tool for technical standardization of the Ki67 immunohistochemical assay. Modern Pathol Off J United States Can Acad Pathol Inc Jul 2021;34(7). https://doi.org/10.1038/s41379-021-00745-6.
6. Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AMB, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. Histopathology Aug 2018;73(2):327–338. https://doi.org/10.1111/his.13516.
7. Bankhead P, Fernández JA, McArt DG, Boyle DP, Li G, Loughrey MB, et al. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. Lab Invest 2018;98(1):15–26. https://doi.org/10.1038/labinvest.2017.131.01.
8. Volynskaya Z, Mete O, Pakbaz S, Al-Ghamdi D. Ki67 quantitative interpretation: insights using Image Analysis. J Pathol Inform 2019;10:8. https://doi.org/10.4103/jpi.jpi_76_18.
9. Boukhar S, Gosse M, Bellizzi A, Rajan KDA. Ki-67 proliferation index assessment in gastro-enteropancreatic neuroendocrine tumors by digital image analysis with stringent case and hotspot level concordance requirements. Am J Clin Pathol 2021;156(4). https://doi.org/10.1093/ajcp/aqaa275.09/08/2021.