Validating a semi-quantitative method for assessment of degree of methylene blue staining in sentinel lymph nodes

Abstract

Purpose: To develop a digital algorithm and validate a semi-quantitative scoring method for surface methylene blue (MB) staining in whole lymph nodes (LN).
Methods: Lymph nodes from canine models undergoing sentinel lymph node (SLN) mapping were prospectively assessed *ex vivo* and photographed. Two blinded observers evaluated all images and assigned a score based on surface staining (0 – no blue stain, 1 – 1-50% stained, 2 – 51-100% stained). A standard reference for degree of blue staining was based on signal-to-background ratios using computer-based imaging software with an output measurement of percentage of staining of the LN. Agreement between observers was assessed using the Kappa coefficient.
Results: 124 lymph nodes were included and demonstrated strong agreement ($K = 0.8007$, $p < 0.0001$) between results of semi-quantitative scoring and image analysis. Also, strong interobserver and intraobserver agreement was observed for the scoring system ($K = 0.8051$, $p < 0.0001$ and $K = 0.9493$, $p < 0.0001$, respectively).
Discussion: Agreement between scoring system and imaging software illustrates a validated method in assessing MB staining, without the need for analysis software. The use of a semi-quantitative scoring system shows promise for a simple, objective assessment of MB staining in surgery and for future study. Lymph nodes can have variable surface colour, which can make assessment of blue staining challenging for novice observers in certain cases. This study describes a digital algorithm for quantitative analysis of blue staining in LN thereby providing a novel and objective reporting mechanism in scientific research involving SLN mapping.

1 Introduction

Sentinel lymph node (SLN) mapping is an important prognostic factor for staging various types of cancers to determine prognosis and treatment recommendations[1,2]. Detecting SLNs is achieved by using lymphatic tracers, most commonly injected peritumorally, that delineate lymphatic tracts to the sentinel nodes[3]. The gold standard in SLN mapping employs dual tracer techniques involving a combination of radioisotopes, blue dyes and/or fluorescence to increase reliability[4–6]. However, in specific countries and facilities, methylene blue dye is used alone for SLN mapping in light of its cost effectiveness, accessibility and safe outcomes[4,7,8]. Methylene blue is a non-specific blue dye with a good safety profile and has been described for use as an alternative to isosulfan blue and patent blue dye[2,4,9]. Breast cancer studies that employ methylene blue dye alone suggest comparable lymphatic uptake and results to other blue dyes[2,10–14]. One of the challenges associated with the use of methylene blue in SLN mapping, can be the correct identification of methylene blue staining compared to normal surface staining. The discernment of blue stain on a lymph node becomes difficult when clinicians cannot identify whether the discolouration is due to staining or to natural lymph node tissue pigmentations, since lymph nodes are often heterogenous in morphology[15] and brown tissue can appear as blue. In cases where a dyed lymphatic is not visible, this challenge can post some difficulty to the clinician in confirming that a lymph node is truly sentinel. Ideally, a digital algorithm would be used to

depict true staining of tissue in each clinical case, however this is not practical or efficient. Therefore, an objective, accessible method for methylene blue stain quantification on whole tissue is needed to improve reporting. This study describes the use of a semi-quantitative scoring system and image analysis process for methylene blue stained whole lymph node tissue. The goal is to describe the use of an open-access program for image analysis and validate the semi-quantitative scoring system.

2 Methods

2.1 *Imaging of Lymph Nodes Stained with Methylene Blue*

Lymph nodes were obtained consecutively from canine patients undergoing full regional lymph node extirpation as part of another SLN biopsy study at the Ontario Veterinary College Health Sciences Centre from 2017-2019. In all patients, peritumoral or intratumoral methylene blue was injected at a concentration of 0.5mg/mL and routine lymph node extirpation performed. Lymph nodes were imaged in either an unstandardized or standardized fashion, depending on which part of the accrual period they were removed. For the unstandardized group, imaging of lymph nodes was performed with an iPhone 6 or newer model, equipped with a 12 mega-pixel camera (Apple, California, USA) in an unstandardized fashion with no lighting, background or camera distance controls. For the standardized group, lymph nodes were placed on a uniform white background in a photo lightbox (Amazon, Canada) equipped with LED lights to provide optimal imaging conditions for gross specimens[16,17]. Using an iPhone X equipped with a 12-megapixel wide angle camera and secondary telephoto lens (Apple, California, USA) images were taken at a distance of 20cm from the specimen through a 1cm hole at the top of the box to improve focus of the smartphone lens.

2.2 *Semi-quantitative Scoring of Methylene Blue Stain*

All lymph nodes were assessed as positive or negative for methylene blue staining by the PI in situ and immediately following removal and assigned a score based on the coloration of the surface of the lymph node. After data collection was completed, all images for both the unstandardized and standardized groups were randomized for evaluation. The randomized images were evaluated by 2 investigators (M.O., A.R.) for blue stain visualized on the surface of the lymph node. Evaluation consisted of a scoring system based on the amount of methylene blue. The scores were determined as follows: 0 = no blue stain, 1+ = 1-50% of the surface of the lymph node is stained, 2+ = 51-100% of the lymph node is stained.

2.3
Verification of Image Analysis Method

The image analysis process was verified by testing true negative and true positive lymph node control images. The negative controls are images of clinically normal lymph nodes from routine biopsies, while the true positive lymph nodes are obtained from cases that show distinct and large amounts of blue stain on the surface of the lymph node, which will be further cropped to areas of stain only; in the form of a region-of-interest (ROI)[18]. Due to the use of a pre-existing, built-in plugin for deconvolution of stains and pigments in an image, the plugin contains verified

vectors that correspond to specific stains[19]. The main validation required is if the plugin can detect methylene blue staining on whole tissue specimens and the threshold that disseminates dark blue staining from false-positive signals. Based on the output of the plugin, negative control images were tested to adjust the threshold of "blue" staining[20]. The threshold depicts that dark pixels corresponding to "blue" range from 0 (being the darkest) to 125. Pixels above this threshold were observed to be false signals of blue due to image reflection, image quality or dark tissue. Using a threshold of 0-125 after running the deconvolution plug-in on negative control images depicted 0% of stain detected for all controls. Similarly, positive control pictures that primarily focused on heavily stained regions shows >95% signal detection.

Quantification of Methylene Blue Stain Using Image Analysis

Images were analyzed in 6 randomized groups of 25; consisting of 124 images analyzed in total. Images were processed and analyzed in FIJI (National Institutes of Health, Bethesda, Maryland, USA), a distribution of an open-source image processing program (ImageJ)[19]. Colour and background corrections were performed on the true colour (RGB) image using the "Subtract Background" feature and auto adjustment function of the "Brightness/Contrast" tool. The scale was set to "no scale" and, by default, the area was measured in pixels. A ROI was drawn around the whole lymph node using the "Colour Threshold" function, producing an automatic threshold over the image that was measured as the area of the entire surface of the lymph node. If auto threshold did not accurately differentiate the area of the lymph node from the image background, the thresholding brightness and saturation levels were adjusted. Subsequent use of the Colour Deconvolution plugin on the RGB image separated pigments into channels. This plugin is usually applied for the purpose of separating multiple histological stains in a tissue sample[21]. The plugin produces a choice of vectors which are associated with specific dye mixtures. For the purpose of this study, the Gimesa vector was chosen. Gimesa is a dye that contains a mixture of methylene blue, eosin, and an optional third component, such as Azure B[21]. Even though the Gimesa setting is for a combination of three stains, the methylene blue vector is verified and readily isolated[21]. The output of the other channels will not show meaningful signals for stains, such as eosin. Once the plugin runs the command, it produces 3 channels of the image in 8-bit grayscale format. For the purpose of this study, the "Colour_1" window was used since it corresponds to the methylene blue channel. The 8-bit image underwent thresholding (Threshold function) using a threshold of 0-125 and a binary image of the thresholded area was generated. This is a strict threshold based on controls to only detect dark pixels. This threshold was measured and generated the area of methylene blue stain on the surface of the lymph node. The amount of methylene blue stain on the surface of the node is calculated by:

$$\frac{Area\ of\ MB\ stain}{Area\ of\ lymph\ node} \times 100.$$

2.4 *Statistical Analysis*

All analyses were completed using SAS 9.3 by a biostatistician (G.M.) who was blinded to the clinical procedures and assessment protocols. Descriptive analysis was performed with summaries of frequencies and agreement percentages in contingency tables. Weighted kappa (κ) statistics were used to assess agreement of assessment modalities, inter- and intra-rater agreement, and agreement between scoring settings. Weighted kappa values with 95%

confidence intervals (CI) were calculated to determine strength of agreement. Coefficients in the range of 0.21–0.40 were interpreted as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1 as almost perfect agreement[22]. Statistical significance was set at a two-sided p-value less than 0.05. The sample size is justified using the weighted kappa and degree of agreement.

Results

A total of 120 lymph nodes were collected from 29 clinical cases of patients undergoing lymph node extirpation with SLN mapping. Eight control lymph node images from 3 clinical cases were also included for analysis. Four images were excluded due to poor image quality or small size of the lymph node resulting in a sample size of 124. Lymph node images were sorted and analysed as described in the methods (Figure 1a). A total of 89 images were included in the unstandardized group and 31 images in the standardized group. The majority of lymph nodes (66.1%) were considered negative for blue staining based on visual assessment, with 29% and 4.8% falling into the 1+ and 2+ categories, respectively. Image analysis was successfully performed and documented in all lymph node images using ImageJ. There was significant concordance between the scoring system and image analysis ($\kappa = 0.8007$ [0.70–0.90], $P < 0.0001$) (Table 1). Interobserver and intraobserver agreement of scoring was strong (($\kappa = 0.8051$ [71–89], $P < 0.0001$ and $\kappa = 0.9493$ [0.89–1], $P < 0.0001$, respectively). Rater 1 and rater 2 scoring compared to the gold standard displayed similar and substantial agreement ($\kappa = 0.7915$, $P < 0.0001$ and $\kappa = 0.8098$, $P < 0.0001$, respectively). Evaluation of settings based on *ex vivo* scoring and scoring based on images displayed almost perfect agreement ($\kappa = 0.9215$ [0.86–0.98], $P < 0.0001$) (Figure 2).

Discussion

Sentinel lymph node mapping relies on accurate visual identification of lymph nodes thought to be "sentinel" based on tracer uptake and stains. Literature and research methodologies lack reporting and objective classification of lymph node staining obtained from SLN biopsies. Reports indicate identification rates (IR) as a point of success in the SLN mapping process[4], although appropriate criteria for successful identification of a SLN is often missing. Clinical trials involving SLN mapping outline inclusion criteria for SLNs reported based on nodes that are only stained blue[2,8,23–28], blue and non-blue nodes with dye uptake in afferent lymphatic channels[11,12,14,29–35], or do not comment on IR inclusion criteria[36–38]. The inconsistent or lack of standardized, objective reporting across SLN mapping trials and cases that utilize methylene blue skews accuracy and reduces comparability of results between studies. Improving the reporting process of identified SLNs positive for methylene blue staining can influence the results of studies and enhance discernment of clinicians that assess these nodes.

In this report, we propose a simple, semi-quantitative scoring system to clinically assess methylene blue staining on lymph nodes extirpated during SLN mapping and validated the system using an image analysis process developed for quantification of stain. The assessment of staining is based on the amount of stain present on the surface of lymph node tissue. Our data shows a strong agreement between the scoring system and image analysis that was statistically significant. Immunohistochemistry (IHC) is the gold standard for evaluating stains, however this

method is not clinically feasible for postoperative evaluation and is not developed to detect intraoperative methylene blue staining[39]. In contrast, image analysis provides an objective, quick examination of tissue and can be used to improve histology process[40–44]. The image analysis method demonstrated in this study was verified using images of negative control lymph nodes that did not pick up signals for stain and visibly true positive nodes that quantified all the methylene blue staining (Figure 1b). During image analysis, the colour deconvolution (CD) method was employed due to the heterogeneity of lymph node tissue[15,45]. Colour deconvolution allows the separation of RGB colours from images into stain channels made with specific vectors[46,47]. These stain channels are in grayscale and correspond to the intensity of a particular stain found in the image[46]. This analysis plugin determines the density of stain in areas where multiples stains are co-localized[41]. Other studies, such as one by Onder *et al.* (2014) have evaluated the robustness of the CD process and found that CD displayed significantly higher sensitivity in classification of stained samples without compromising specificity when compared to hue-saturation-intensity (HSI) separation method[47]. Due to the superiority of CD in being able to detect dark areas that correspond to brown or blue which HSI could not differentiate, this plugin was employed in our image analysis. The three-grade scoring system developed in this study provides a simple semi-quantitative and accurate assessment of lymph nodes in a clinical setting and is consistent with existing scoring processes that are based on the overall stain intensity (i.e. percentage of cells stained)[40,42,48]. The validation of our scoring system is based on IHC scoring methods that use image analysis to objectively validate the visual assessment[49–51].

Our scoring system has substantial interobserver agreement depicting there is little variance between scores given by rater 1 and rater 2. The lack of variability between two scorers illustrates that this visual assessment can be done by different observers and still yield the same score given to a sample. Variability in scores between observers is seen when lymph node tissue colouration is ambiguous or if image quality is poor. The intraobserver agreement of scores in our study are near perfect and scores did not even vary when the blinded rater scored the randomized images in batches at a different time. Scores of different observers compared to the gold standard result were in strong agreement, further showing that there is little variability between the score determined by individuals and that of the gold standard. Also, in-person scores of *ex vivo* SLNs were in agreement with scores based on images. This result displays the congruency between samples and images of the samples, where colour and quality are not misconstrued allowing for scoring to take place at any time based on available pictures.

A strength of this study is the high degree of agreement that allows for the sample size to be sufficient with 100% power. The image analysis program is practical and can easily be utilized by clinicians, researchers, and other hospital or lab staff since FIJI is an opensource, free software tailored for biosciences[19]. The accessibility of the program can allow widespread use, allowing for limited variability in analysis programs between studies. Researchers may be less inclined to adopt this method of analysis if an expensive specialized software was required. Also, this visual assessment and image analysis program provides a short learning curve in appropriately scoring and analyzing the lymph nodes. A limitation of this study was the distribution of samples represented in each scoring category. In the concurrent study, patients were undergoing total lymph node basin extirpation to assess the accuracy of SLN mapping in that patient population. As a result, there are a large number of lymph node samples imaged that were negative for methylene blue stain (66.1%). Since the goal of this study was validation, we

felt it was important to include this population since it was possible that our visual assessment of a lymph node negative for staining could be different than what was found with image analysis. As a result, our study has an unequal sample size within score categories of 1+ and 2+ (33.8%). However, this does not affect the agreement coefficients due to the analysis and scoring system having low false negative and false positive rates (Table 1). Another potential limitation is the use of CD. Color devolution can have pitfalls in its ability to detect dark areas, where brown pigment can be falsely recognized as dark blue pixels[18,46], usually when stains like diaminobenzidine (DAB) are used. This was seen in our results where SLNs scored as 0 are detected as a category of 1+ by image analysis (8.87%) since dark tissue is recognized as traces of blue. During our analysis this was not a prominent issue due to the vector used and the strict thresholds. If the program does falsely recognize dark tissue as blue stain, the detection percentage is low and clinically negligible to the human eye where it ranges from 1-2% of stain detected. A final perceived limitation of the study may be the lack of automation to further the objectivity of the analysis process. An automated process for determining ROIs would be efficient and robust, however lymph nodes vary greatly in size and shape which make it difficult to tailor specific macros for ROI creation.

In conclusion, we developed a pragmatic visual and analytic assessment system to evaluate the degree of blue staining in extirpated lymph nodes when SLN mapping is performed using methylene blue dye. The scoring system and quantitative image analysis program have strong agreement which shows the validity of the visual assessment. This assessment workflow allows for standardized reporting of clinical research to improve comparability and consistency of results in SLN mapping of various cancers utilizing methylene blue. The validated visual scoring system provides an accessible and objective measure in a clinical setting when image analysis is not available. It is not yet known if methylene blue staining patterns and uptake has significance for patterns of metastasis and outcomes, but a scoring and digital quantification system is required to investigate such research and is an objective of future directions.

Acknowledgements

We would like to thank Gabrielle Monteith (University of Guelph) for assisting with statistics used in this report.

Disclosure of Potential Conflicts of Interest

The authors have no conflicts of interest to disclose.

References

1. Somasundaram SK, Chicken DW, Keshtgar MRS. Detection of the sentinel lymph node in breast cancer. *Br Med Bull*. 2007;84(1):117-131. doi:10.1093/bmb/ldm032

2. Mathelin C, Croce S, Brasse D, et al. Methylene blue dye, an accurate dye for sentinel

lymph node identification in early breast cancer. *Anticancer Res*. 2009;29(10):4119-4125. http://www.ncbi.nlm.nih.gov/pubmed/19846959. Accessed July 17, 2019.

3. Schauer AJ, Becker W, Reiser M, Possinger K. Definition of the Sentinel Lymph Node and Basic Principles of Detection. In: *The Sentinel Lymph Node Concept*. Berlin/Heidelberg: Springer-Verlag; 2005:3-4. doi:10.1007/3-540-26393-4_1
4. Li J, Chen X, Qi M, Li Y. Sentinel lymph node biopsy mapped with methylene blue dye alone in patients with breast cancer: A systematic review and meta-analysis. Metze K, ed. *PLoS One*. 2018;13(9):e0204364. doi:10.1371/journal.pone.0204364
5. Guo J, Yang H, Wang S, et al. Comparison of sentinel lymph node biopsy guided by indocyanine green, blue dye, and their combination in breast cancer patients: a prospective cohort study. *World J Surg Oncol*. 2017;15(1):196. doi:10.1186/s12957-017-1264-7
6. He P-S, Li F, Li G-H, Guo C, Chen T-J. The combination of blue dye and radioisotope versus radioisotope alone during sentinel lymph node biopsy for breast cancer: a systematic review. *BMC Cancer*. 2016;16(1):107. doi:10.1186/s12885-016-2137-0
7. Vishnoi JR, Kumar V, Gupta S, et al. Outcome of sentinel lymph node biopsy in early-stage squamous cell carcinoma of the oral cavity with methylene blue dye alone: a prospective validation study. *Br J Oral Maxillofac Surg*. 2019;57:755-759. doi:10.1016/j.bjoms.2019.06.025
8. Ramamurthy R, Kottayasamy Seenivasagam R, Shanmugam S, Palanivelu K. A prospective study on sentinel lymph node biopsy in early oral cancers using methylene blue dye alone. *Indian J Surg Oncol*. 2014;5(3):178-183. doi:10.1007/s13193-014-0337-0
9. Tsopelas C, Sutton R. Why certain dyes are useful for localizing the sentinel lymph node. *J Nucl Med*. 2002;43(10):1377-1382. http://www.ncbi.nlm.nih.gov/pubmed/12368377.

Accessed September 18, 2019.

10. Dewachter P, Mouton-Faivre C, Benhaijoub A, Abel-Decollogne F, Mertes PM. Anaphylactic reaction to patent blue V after sentinel lymph node biopsy. *Acta Anaesthesiol Scand*. 2006;50(2):245-247. doi:10.1111/j.1399-6576.2006.00912.x

11. Golshan M, Nakhlis F. Can Methylene Blue Only Be Used in Sentinel Lymph Node Biopsy for Breast Cancer? *Breast J*. 2006;12(5):428-430. doi:10.1111/j.1075-122X.2006.00299.x

12. Simmons R, Thevarajah S, Brennan MB, Christos P, Osborne M. Methylene Blue Dye as an Alternative to Isosulfan Blue Dye for Sentinel Lymph Node Localization. *Ann Surg Oncol*. 2003;10(3):242-247. doi:10.1245/ASO.2003.04.021

13. Varghese P, Abdel-Rahman AT, Akberali S, Mostafa A, Gattuso JM, Carpenter R. Methylene Blue Dye-A Safe and Effective Alternative for Sentinel Lymph Node Localization. *Breast J*. 2008;14(1):61-67. doi:10.1111/j.1524-4741.2007.00519.x

14. Zakaria S, Hoskin TL, Degnim AC. Safety and technical success of methylene blue dye for lymphatic mapping in breast cancer. *Am J Surg*. 2008;196(2):228-233. doi:10.1016/j.amjsurg.2007.08.060

15. Luscieti P, Hubschmid T, Cottier H, Hess MW, Sobin LH. Human lymph node morphology as a function of age and site. *J Clin Pathol*. 1980;33(5):454-461. doi:10.1136/jcp.33.5.454

16. Rampy BA, Glassy EF. Pathology Gross Photography: The Beginning of Digital Pathology. *Clin Lab Med*. 2016;36(1):67-87. doi:10.1016/j.cll.2015.09.008

17. Madrigal E, Prajapati S, Hernandez-Prera JC. Introducing a Virtual Reality Experience in Anatomic Pathology Education. *Am J Clin Pathol Oct*. 2016;146:462-468.

doi:10.1093/AJCP/AQW133

18. Shu J, Dolman GE, Duan J, Qiu G, Ilyas M. Statistical colour models: an automated digital image analysis method for quantification of histological biomarkers. *Biomed Eng Online*. 2016;15(1):46. doi:10.1186/s12938-016-0161-6

19. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682. doi:10.1038/nmeth.2019

20. Lehmann TM, Jo¨ J, Bredno J. The Practice of Informatics Strategies to Configure Image Analysis Algorithms for Clinical Usage. *J Am Med Informatics Assoc*. 2005;12:497-504. doi:10.1197/jamia.M1652

21. Landini G. Colour Deconvolution. http://www.mecourse.com/landinig/software/cdeconv/cdeconv.html. Published 2015. Accessed March 26, 2019.

22. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213-220. doi:10.1037/h0026256

23. Abu-Rustum NR, Khoury-Collado F, Pandit-Taskar N, et al. Sentinel lymph node mapping for grade 1 endometrial cancer: is it the answer to the surgical staging dilemma? *Gynecol Oncol*. 2009;113(2):163-169. doi:10.1016/j.ygyno.2009.01.003

24. Bara T, Gurzu S, Jung I, et al. Sentinel node biospy using intravital blue dye: An useful technique for identification of skip metastases in gastric cancer. *Medicine (Baltimore)*. 2019;98(12):e14951. doi:10.1097/MD.0000000000014951

25. Chagpar AB, Martin RC, Scoggins CR, et al. Factors predicting failure to identify a sentinel lymph node in breast cancer. *Surgery*. 2005;138(1):56-63. doi:10.1016/J.SURG.2005.03.003

26. Coskun G, Dogan L, Karaman N, Ozaslan C, Atalay C. Value of sentinel lymph node biopsy in breast cancer patients with previous excisional biopsy. *J Breast Cancer*. 2012;15(1):87-90. doi:10.4048/jbc.2012.15.1.87

27. Cox CE, Dupont E, Whitehead GF, et al. Age and Body Mass Index May Increase the Chance of Failure in Sentinel Lymph Node Biopsy for Women with Breast Cancer. *Breast J*. 2002;8(2):88-91. doi:10.1046/j.1524-4741.2002.08203.x

28. Nandu V V, Chaudhari MS. Efficacy of Sentinel Lymph Node Biopsy in Detecting Axillary Metastasis in Breast Cancer Using Methylene Blue. *Indian J Surg Oncol*. 2017;8(2):109-112. doi:10.1007/s13193-016-0616-z

29. Brahma B, Putri RI, Karsono R, et al. The predictive value of methylene blue dye as a single technique in breast cancer sentinel node biopsy: a study from Dharmais Cancer Hospital. *World J Surg Oncol*. 2017;15(1):41. doi:10.1186/s12957-017-1113-8

30. Çolakoğlu MK, Güven E, Akgül GG, Doğan L, Gülçelik MA. Biological Subtypes of Breast Cancer and Sentinel Lymph Node Biopsy. *Eur J breast Heal*. 2018;14(2):100-104. doi:10.5152/ejbh.2018.3780

31. Bouquet de Jolinière J, Major A, Khomsi F, Ben Ali N, Guillou L, Feki A. The Sentinel Lymph Node in Breast Cancer: Problems Posed by Examination During Surgery. A Review of Current Literature and Management. *Front Surg*. 2018;5:56. doi:10.3389/fsurg.2018.00056

32. Koller M, Barsuk D, Zippel D, Engelberg S, Ben-Ari G, Papa MZ. *Sentinel Lymph Node Involvement a Predictor for Axillary Node Status with Breast Cancer Has the Time Come?* Vol 24.; 1998. http://www.hbuk.co.uk/. Accessed September 18, 2019.

33. Li H, Jun Z, Zhi-Cheng G, Xiang Q. Factors that affect the false negative rate of sentinel

lymph node mapping with methylene blue dye alone in breast cancer. *J Int Med Res*. September 2019:030006051982741. doi:10.1177/0300060519827413

34. Nour A. Efficacy of Methylene Blue Dye in Localization of Sentinel Lymph Node in Breast Cancer Patients. *Breast J*. 2004;10(5):388-391. doi:10.1111/j.1075-122X.2004.21360.x

35. Povoski SP, Olsen JO, Young DC, et al. Prospective Randomized Clinical Trial Comparing Intradermal, Intraparenchymal, and Subareolar Injection Routes for Sentinel Lymph Node Mapping and Biopsy in Breast Cancer. *Ann Surg Oncol*. 2006;13(11):1412-1421. doi:10.1245/s10434-006-9022-z

36. Djurisic I, Santrac N, Buta M, et al. Can we use frozen section analysis of sentinel lymph nodes mapped with methylene blue dye for decision making upon one-time axillary dissection in breast carcinoma surgery in developing countries? *J BUON*. 2015;20(2):492-497. http://www.ncbi.nlm.nih.gov/pubmed/26011341. Accessed October 19, 2019.

37. Özdemir A, Mayir B, Demirbakan K, Oygür N. Efficacy of Methylene Blue in Sentinel Lymph Node Biopsy for Early Breast Cancer. *J breast Heal*. 2014;10(2):88-91. doi:10.5152/tjbh.2014.1914

38. Restaino S, Ronsini C, Finelli A, Perrone E, Scambia G, Fanfani F. Role of blue dye for sentinel lymph node detection in early endometrial cancer. *Gynecol Surg*. 2017;14(1):23. doi:10.1186/s10397-017-1026-0

39. Walker RA. Quantification of immunohistochemistry-issues concerning methods, utility and semiquantitative assessment I. *Histopathology*. 2006;49(4):406-410. doi:10.1111/j.1365-2559.2006.02514.x

40. Varghese F, Bukhari AB, Malhotra R, De A. IHC Profiler: An Open Source Plugin for the

Quantitative Evaluation and Automated Scoring of Immunohistochemistry Images of Human Tissue Samples. *PLoS One*. 2014;9(5):96801. doi:10.1371/journal.pone.0096801

41. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*. 2001;23(4):291-299. http://www.ncbi.nlm.nih.gov/pubmed/11531144. Accessed March 26, 2019.

42. Choudhury KR, Yagle KJ, Swanson PE, Krohn KA, Rajendran JG. A robust automated measure of average antibody staining in immunohistochemistry images. *J Histochem Cytochem*. 2010;58(2):95-107. doi:10.1369/jhc.2009.953554

43. Brey EM, Lalani Z, Johnston C, et al. Automated Selection of DAB-labeled Tissue for Immunohistochemical Quantification. *J Histochem Cytochem*. 2003;51(5):575-584. doi:10.1177/002215540305100503

44. Vrekoussis T, Chaniotis V, Navrozoglou I, et al. Image analysis of breast cancer immunohistochemistry-stained sections using ImageJ: an RGB-based model. *Anticancer Res*. 2009;29(12):4995-4998. http://www.ncbi.nlm.nih.gov/pubmed/20044607. Accessed December 12, 2019.

45. Ulvmar MH, Mäkinen T. Heterogeneity in the lymphatic vascular system and its origin. *Cardiovasc Res*. 2016;111(4):310-321. doi:10.1093/cvr/cvw175

46. Alsubaie N, Trahearn N, Raza SEA, Snead D, Rajpoot NM. Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation. Bianchi C, ed. *PLoS One*. 2017;12(1):e0169875. doi:10.1371/journal.pone.0169875

47. Onder D, Zengin S, Sarioglu S. A Review on Color Normalization and Color Deconvolution Methods in Histopathology. *Appl Immunohistochem Mol Morphol*. 2014;22(10):713-719. doi:10.1097/PAI.0000000000000003

48. Yagi Y, Gilbertson JR. Digital imaging in pathology: the case for standardization. *J Telemed Telecare*. 2005;11(3):109-116. doi:10.1258/1357633053688705

49. Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R. A Comparison of Visual Assessment and Automated Digital Image Analysis of Ki67 Labeling Index in Breast Cancer. Coppola D, ed. *PLoS One*. 2016;11(2):e0150505. doi:10.1371/journal.pone.0150505

50. Loughrey MB, Bankhead P, Coleman HG, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using *QuPath* digital image analysis. *Histopathology*. 2018;73(2):327-338. doi:10.1111/his.13516

51. Jakobsen MR, Teerapakpinyo C, Shuangshoti S, Keelawat S. Comparison between digital image analysis and visual assessment of immunohistochemical HER2 expression in breast cancer. *Pathol - Res Pract*. 2018;214(12):2087-2092. doi:10.1016/j.prp.2018.10.015

Figures

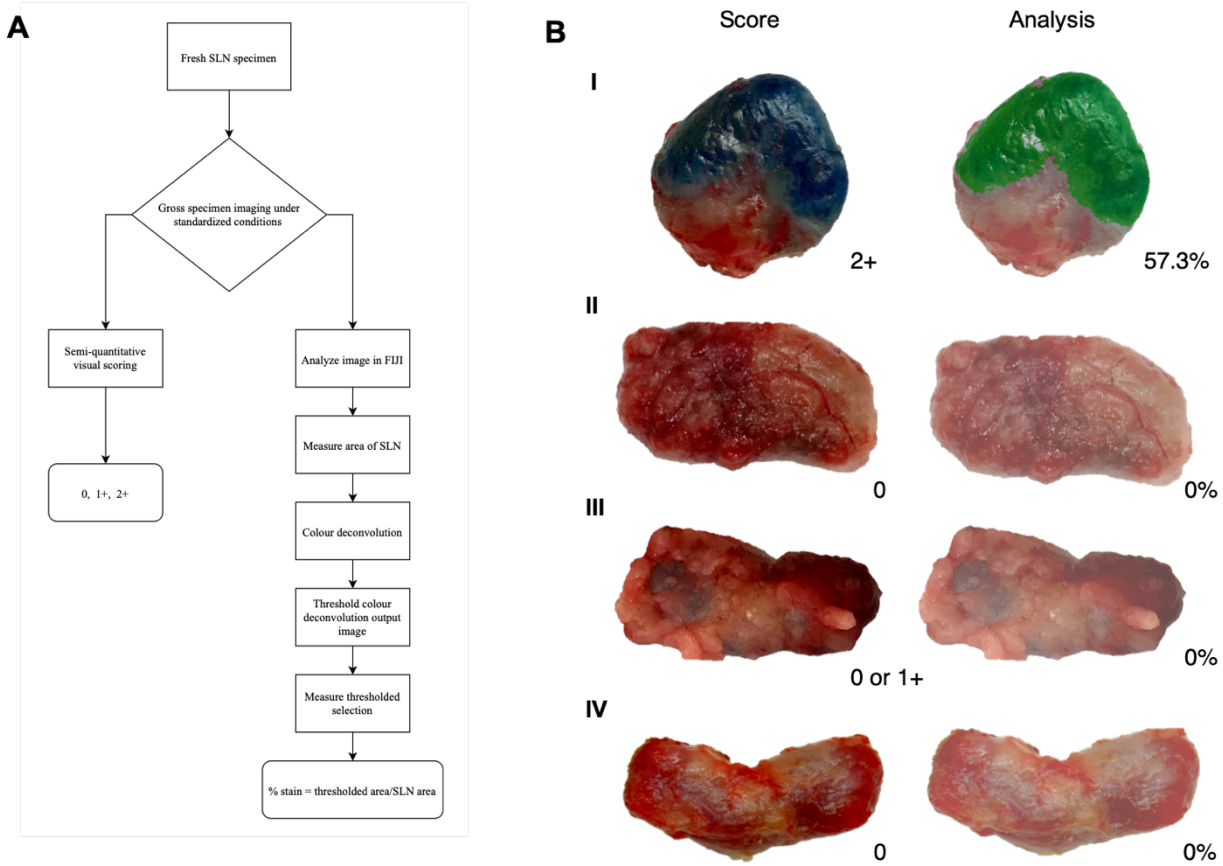

Figure 1 Visual assessment and image analysis outcomes. **a** Workflow of assessment for methylene blue stained lymph nodes. **b** Score and analysis of SLNs. Depicting a positive node (I), a negative node (II), an ambiguous node (III), and a negative control node (IV); all with agreement between score and analysis.

Table 1 Frequency of agreement between visual scoring and analysis

		Quantitative image analysis			Total
		0	1+	2+	
Semi-quantitative scoring system	0	82 (66.13%)	11 (8.87%)	0 (0)	93 (75%)
	1+	0 (0%)	23 (18.55%)	0 (0%)	23 (18.55%)
	2+	0 (0%)	2 (1.61%)	6 (4.84%)	8 (6.45%)
	Total	82 (66.13%)	36 (29.03%)	6 (4.84%)	124 (100.00%)
Percentage agreement	89.5%				
Weighted kappa	0.8007 [0.70 – 0.90]				

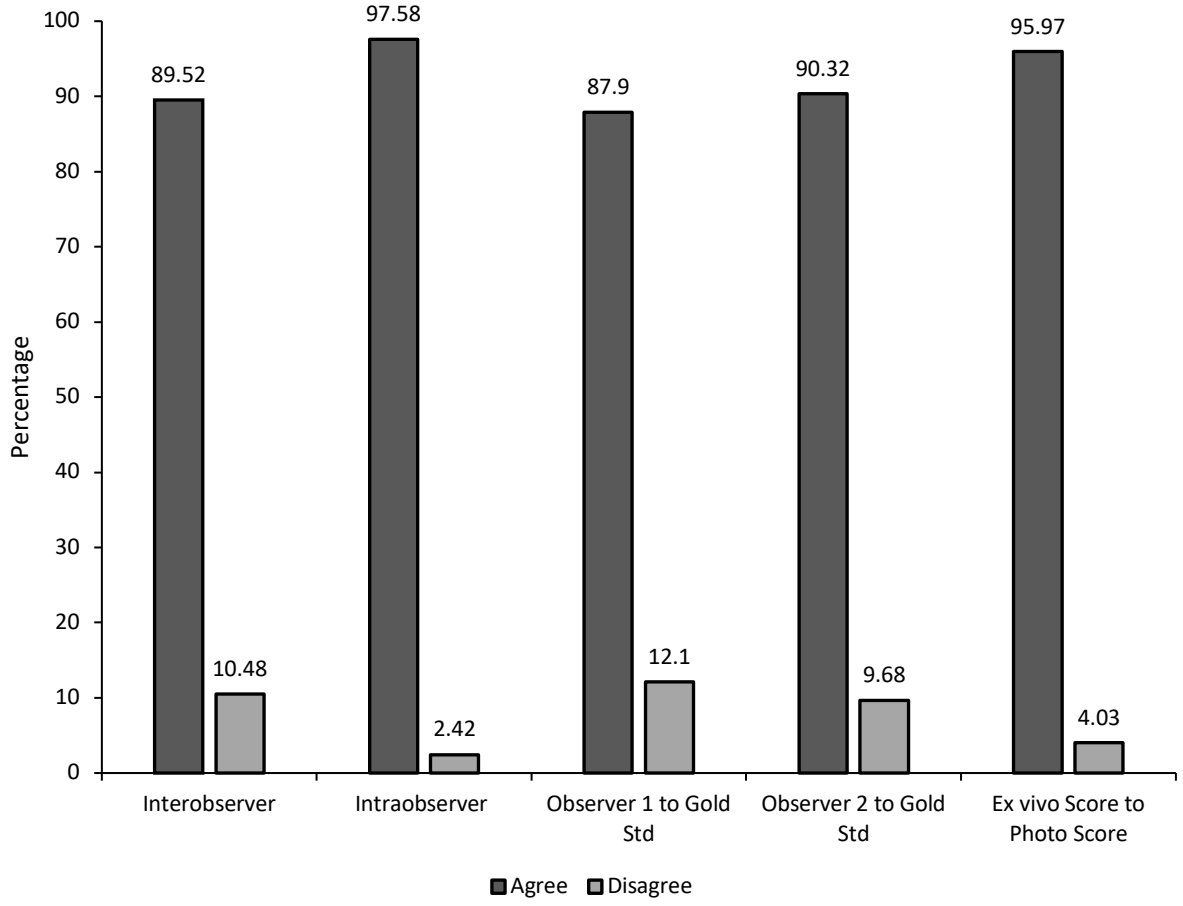

Figure 2 Histogram showing the frequencies of agreement and disagreement of scores in interobserver test, interobserver test, comparing observers to the gold standard, and evaluating observer settings. All frequencies are statistically significant ($P < 0.0001$).