OPEN

Tissue Thickness Effects on Immunohistochemical Staining Intensity of Markers of Cancer

Adrienne S. McCampbell, PhD, Varun Raghunathan, PhD,* May Tom-Moy, PhD,* Richard K. Workman, PhD,* Rick Haven, PhD,* Amir Ben-Dor, PhD,† Ole F. Rasmussen, PhD,‡ Lars Jacobsen, PhD,‡ Martin Lindberg, PhD,‡ N. Alice Yamada, PhD,* and Carol Schembri**

Abstract: High-quality patient samples are required for reliable immunohistochemistry test outcomes that provide a significant benefit for patient care. Among the preanalytic variables in tissue handling, tissue thickness is thought to be easily controlled; however, whether the thickness of the tissue effects the staining intensity for antibody immunohistochemistry has not been quantitatively demonstrated. To investigate, we cut multiblock tissue sections of tonsil, liver, and kidney at 2, 4, 6, and 8 μm thicknesses. Interferometry measurements of the sectioned paraffin showed a <1 μm variation within a preset microtome thickness. Sections were then immunostained with antibodies targeting different cellular localizations; Ki-67 and BCL6 (nuclear), CD7 (membranous), and cytokeratin (cytoplasmic). A pathologist annotated regions of interest for each marker and performed brightfield and whole-slide visual scoring. Then a pixel-wise processing algorithm determined intensity of each pixel in these regions of interest and binned them into predetermined 0, 1+, 2+, or 3+ intensities. Visual scores from brightfield and whole-slide images were highly correlated to the percentage of pixels in each intensity bin. A stepwise increase was observed in pathologist scores and algorithmically defined percentage of pixels in each bin with increasing thickness demonstrating that changes in preset section thickness impacts staining intensity. The use of tissue thickness outside vendors' recommendations might change the intensity including the proportion of positive and negative cells and eventually the overall diagnosis outcome. Therefore, we recommend that tissue be consistently cut within the middle of thickness range specified by the assay manufacturer.

Received for publication June 5, 2017; accepted August 16, 2017.

From the *Agilent Research Laboratories, Agilent Technologies, Santa Clara, CA; †Agilent Research Laboratories, Agilent Technologies, Tel Aviv, Israel; and ‡Agilent Technologies, Glostrup, Denmark.

The authors declare no conflict of interest.

Reprints: Adrienne S. McCampbell, PhD, Agilent Research Laboratories, Agilent Technologies Inc., 5301 Stevens Creek Blvd., Santa Clara, CA 95051 (e-mail: adrienne.mccampbell@agilent.com).

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.appliedimmunohist.com.

Key Words: tissue thickness, staining intensity, Ki-67, BCL6, CD7, CK, algorithm

(*Appl Immunohistochem Mol Morphol* 2017;00:000–000)

To enable high quality in the diagnosis of cancer patients, there is a strong need for high-quality patient sample material. This need is further strengthened with increased focus on personalized medicine where the outcome of a single test can dictate treatment approaches. Many cancer treatment approaches rely upon tissue markers that are indicative of proliferation or are specific to a drugtarget and inform a corresponding therapy and thus rely on immunohistochemistry (IHC) as the standard test method. For example, IHC testing for the estrogen receptor and progesterone receptor, which is crucial for breast cancer subtype diagnosis and therapy, must be determined for all invasive breast cancer recurrences,[1] thus it is critical that IHC tests are accurate. According to Hammond and colleagues, however, up to 20% of IHC determinations for estrogen receptor and progesterone receptor status, are inaccurate due to preanalytical and analytical variables that affect thresholds for positivity and interpretation of scoring. Standardization of IHC and molecular testing is crucial to ensure correct test results avoiding false positive and false negatives; furthermore, the implementation of metrics and guidance for enabling high-quality biospecimens for testing is recognized to be of utmost importance in light of the use of predictive and pharmacodynamic markers in personalized medicine.[2]

Immunohistologic variables that may be subject to standardization include preanalytical and analytic variables. Preanalytic variables include the conditions during tissue acquisition such as warm and cold ischemic times during surgery or autopsy, and fixative type and duration of exposure before paraffin embedding. Analytic variables include the antigen retrieval conditions, inclusion of appropriate control tissues and section thickness.[3] In the United States, the standardization efforts are led by the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) whose Testing Guideline Steering Committee pathologist members continually update published templates for the interpretation and reporting of oncology biomarkers.[4] Current ASCO/CAP guidelines for handling surgical pathology

specimens call for routine tissue types to be cut at 4 to 5 μm in thickness at microtomy.[5]

Attempts have been made to standardize methodology through comprehensive recommendations on preanalytical and analytical assessment, and interpretation and scoring of Ki-67, a strong prognostic indicator of breast, colorectal, ovarian, and other cancers, by the International Ki-67 in Breast Cancer Working Group.[6–10] Preanalytic issues that were considered to adversely affect Ki-67 measurement included the type of biopsy, type of fixative, time of fixation, specimen storage and the proper formalin fixation, and paraffin embedding technique. However, the thickness of the cut tissue sections was not a consideration for their recommended methodology. Indeed, a subsequent Ki-67 reproducibility study whose aims were to assess the Working Groups set guidelines for analysis and reporting identified substantial variability in Ki-67 scoring among some of the world's most experienced laboratories.[11] Intralaboratory reproducibility was high with an interclass correlation coefficient of 0.94 which was reduced to 0.71 for interlaboratory reproducibility with contributing factors including subjective assessment of staining positivity. Polley et al[11] reported a geometric mean of Ki-67 values for each laboratory across 100 cases that ranged from 7.1% to 23.9% indicating that Ki-67 values and cutoffs for clinical decision-making cannot be transferred between laboratories.

To investigate the impact of preset section thickness on staining intensity, we cut formalin-fixed, paraffin-embedded (FFPE) multitissue blocks at 4 preset section thicknesses. We used interferometry, a technique that uses patterns of interference created by superposition of ≥ 2 waves[12] to measure variations in the thickness of paraffin sections cut at preset thicknesses to verify the thickness. We then interrogated differences in staining intensity versus thickness by comparing the pathologist's visual score in brightfield and from whole-slide images (WSIs) and the stain intensity of those WSIs using a quantitative imaging approach.

METHODS

Tissues

Tissue multiblocks including tonsil, liver, and kidney with tissue fixed from 6 to 48 hours were prepared according to ASCO/CAP guidelines. The tissue used were residual tissue specimens from individual patients whose identities are not traceable. Studies were performed in accordance with the Declaration of Helsinki and Good Clinical Practice. No human subjects were directly involved in this study and therefore, no Institutional Review Board ethics committees' approval was required. Diagnosis was confirmed by a pathologist for a diagnosis of tonsil with hyperplasia, normal liver, and normal kidney.

For determination of repeatability in the thickness of cuts and staining intensity 5 serial sections of FFPE multiblock were cut at 6 μm thickness using a microtome (Leica Biosystems, Buffalo Grove, IL) and mounted onto FLEX IHC Microscope slides (Agilent Technologies Inc., Santa Clara, California). For determination of the effect of tissue thickness on staining intensity, FFPE multiblock was cut at 2, 4, 6, or 8 μm thickness cuts and mounted onto FLEX IHC Microscope Slides (Agilent Technologies Inc.). All tissue sections were cut consecutively to best facilitate direct comparison across the sections. Hematoxylin and eosin–stained slides were microscopically evaluated by a pathologist to verify the presence of histologically normal liver and kidney and to confirm the diagnosis of hyperplasia in the tonsil.

Antibodies and Reagents

All Dako antibodies were obtained from Agilent Technologies Inc. The antibodies were: FLEX Monoclonal Mouse Anti-Human BCL6 Protein, Clone PG-B6p, Ready-to-use (Link) (code IR625), Monoclonal Mouse Anti-Human Cytokeratin, Clone AE1/AE3 Ready-to-use (Link) (code IR053), Monoclonal Mouse Anti-Human CD7, Clone CBC.37 Ready-to-use (Link) (code IR643), and Monoclonal Mouse Anti-Human Ki-67 Antigen, Clone MIB-1 Ready-to-use (Link) (code IR626). The antibodies for BCL6 and cytokeratin (CK) were used with the EnVision FLEX, High pH visualization system (Link) (code K8000) (Agilent Technologies Inc.). The antibodies for CD7 and Ki-67 were used with the EnVision FLEX, Low pH visualization system (Link) (code K8005) (Agilent Technologies Inc.) according to manufacturer's instructions for use. Autostainer Classic with upgraded software for the Autostainer Link 48 automated staining platform was used for IHC with Dako PT Link (Agilent Technologies Inc.) being used for pretreatment.

Interferometer Measurements

To determine the range of thickness in the manually cut paraffin-embedded sections, we sectioned paraffin blocks using a Leica microtome at the preset 2, 4, 6, and 8 μm thickness cuts ($n = 50$ per thickness) and measured the paraffin thickness using a ZygoZeGage Interferometer with Zygo MX 6.3.0.4 software (Zygo Corporation, Middlefield, CT) along the entire length of the boundary of the paraffin-glass interface and averaged the results. The SD was computed for each slide and then averaged for the overall SD.

IHC

For IHC, all reagents were obtained from Agilent Technologies Inc. FFPE tissue sections were deparaffinized using xylene and rehydrated in a series of graded ethanol washes to 1X PBS. Antigen retrieval was performed using Target Retrieval Solution High pH or Low pH according to Dako PT Link 100 (code PT10027) and the manufacturer's recommended protocol. The immunostaining was performed on a Dako Autostainer classic (Agilent Technologies Inc.). The protocol was programmed as follows: antigen retrieval was performed in either high or low pH, the tissues sections were blocked with Dual Endogeneous Blocking Reagent (code S2003) for 5 minutes, and then the appropriate antibodies in a ready-to-use format were incubated on the tissue sections for 60 minutes. After 2 brief rinses in Wash Buffer

(code K8007), tissue sections were incubated with Polymer Horseradish Peroxidase (code K4063) for 30 minutes, followed by 2 short rinses using Wash Buffer and then substrate-chromogen reagent (diaminobenzidine) was applied for 15 minutes. The sections were rinsed and the slides were coverslipped using Faramount Mounting Medium (code S3025). The sections were not counterstained to allow maximal interpretation of staining intensity with a single chromogen (diaminobenzidine). The negative reagent control(s) that lacked primary antibody were incubated in wash buffer. We immunostained 1 multiblock section per thickness per antibody. Immunostained sections were then scanned to produce a WSI of each single-stained slide. Original immunostained slides were scored for expression of each individual antibody by both bright field microscopy and the WSI system by a pathologist.

Antibody expression for Ki-67, CK, CD7, and BCL6 was scored quantitatively according to the intensity of staining on a scale of 0 (no staining) to 3+ (strong staining). Each antibody was scored by assessing the number of positive cells or nuclei per 100 cells in a region of interest (ROI) for the target tissue. Both Ki-67 and BCL6 expression was scored in nuclei of B cells in the germinal centers of the tonsil tissue. CD7 expression was scored in the membrane of T-cell regions in mucosa-associated lymphoid tissue (MALT) that surrounds the germinal centers in the tonsil tissue. CK expression, which is predominantly cytoplasmic but can also be membranous was scored in the cytoplasm and membranes of hepatocytes in the liver tissue. The ROIs contained 1+, 2+, or 3+ intensity in individual cells or individual nuclei.

Pixel-wise Data Processing of WSI

Pixel-wise processing on the image data were performed. WSIs were scanned for each single-stained slide (n = 16). The WSI was then carefully annotated by a pathologist to select ROIs that were similar in each tissue type at the different precut thickness to prevent errors in the subsequent algorithm determination of the intensities. For Ki-67 and BCL6 germinal centers within tonsil tissue were included in the annotation of the ROIs and the surrounding epithelial cell compartments and MALT were excluded from the ROIs. For CD7, regions of MALT in the tonsil tissue were included in the ROI annotation and the surrounding germinal centers and epithelial cell compartments were excluded. For CK, regions of solid hepatocytes were included in the annotated ROIs with exclusion of the portal triad and surrounding capsule.

ROIs were then used for visual scoring determination and for pixel-wise processing of image data as described. Using the select tissue regions from the scanned image (or ROIs) the red, green, blue values were extracted. The red, green, blue values were then converted to stain absorbance (Absorbance = −log10(pixel value/255)).Outlier pixels were removed with a simple principal component analysis fit. A scoring metric was defined as an average and SD of the absorbance value at the intensity cutoffs corresponding to 0 intensity, 1+ intensity, 2+ intensity, and 3+ intensity as determined visually by a pathologist and the average and SD of the percentage of pixels were calculated for each intensity scoring window or bin. The pathologist visual score of the ROIs and percentage of pixels stained for each marker and each thickness for the appropriate tissue were then correlated.

Statistical Analysis

For continuous variables, the difference between groups was examined by unpaired Student *t* tests or analysis of variance. Statistical significance was defined as $P < 0.05$.

RESULTS

Interferometer Measurements of Sectioned Paraffin

To determine the variability in the thickness of the microtome cut sections in a quantitative manner, we measured the effective thickness by interferometry (Fig. 1) for paraffin surrounding the tissue in multiblocks (n = 50) cut at 2, 4, 6, or 8 μm according to the Leica microtome settings. We specifically measured the average paraffin thickness along the boundary of the paraffin-glass interface that did not contain any embedded tissue. The mean and (SD) of the 2, 4, 6, or 8 μm thickness sections was 2.38 (±0.18), 4.51 (±0.36), 6.75 (±0.50), and 8.70 (±0.65) μm, respectively. The edge thickness of the sectioned paraffin had a total range of <1 μm in thickness around a given section.

Method for Determination of Differences in Staining Intensity With Variations in Tissue Thickness

To compare data obtained from a pathologist visual scores to data obtained from the percentage of pixels stained

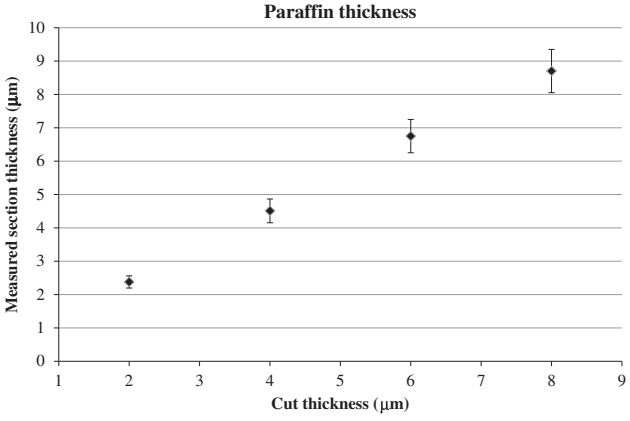

FIGURE 1. Determination of the thickness of paraffin cut by microtomy using interferometry. Paraffin blocks were cut using a Leica microtome programmed to cut 2, 4, 6, or 8 μm sections (n = 50 at each thickness). The thickness of the cut paraffin was then measured using a Zygo interferometer along the boundary of the paraffin-glass interface. Values are interferometer determined mean (μm) (±SD). Cut values are the microtome defined cut thickness.

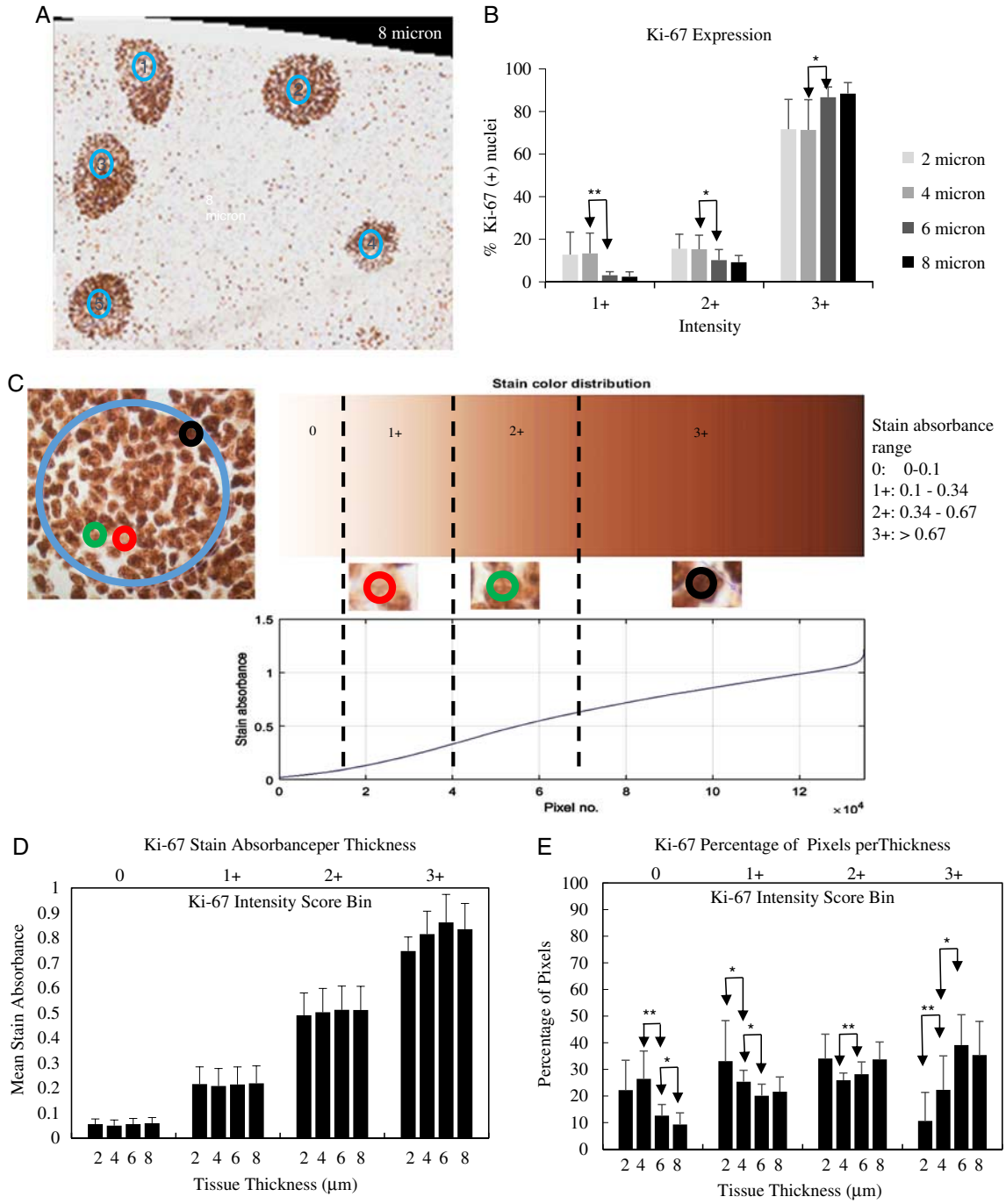

FIGURE 2. Method for determination of differences in staining intensity with tissue thickness. A pathologist scored immunostained sections manually using brightfield microscopy. Immunostained sections were then scanned and the whole-slide image (WSI) was annotated by a pathologist. Regions of interest (ROIs) were selected that represent 100 cells (A; blue circles). The tissue sections were scored visually with nuclei counted that showed a 1+, 2+, or 3+ intensity for multiple ROI's containing 100 B cells each in germinal centers of tonsil tissue for tissue sections cut at 2, 4, 6, or 8 µm (B). Data shown are mean values (± SD). The WSI was annotated for ROI's and used to determine the stain absorbance range for all ROIs selected (C). The stain absorbance was used to determine a stain color distribution. Using the ROIs cutoffs for 1+ (red circle), 2+ (green circle), 3+ (black circle) intensity bins were determined on the stain color distribution map for the algorithmic analysis. The mean stain absorbance (D) and the percentage of pixels (E) in each cutoff window were determined by algorithmic analysis. Visual scores of WSI and the algorithmically determined percentage of pixels were then correlated as a measure of agreement between the 2 methods. This method was repeated for each antibody (CD7, BCL6, or CK) and tissue thickness. *Statistically significant at the $P < 0.05$ level. **Statistically significant at the $P < 0.025$ level.

in an area of a WSI, we developed an algorithm based on pixel-wise processing for image analysis (Fig. 2). A pathologist scored the staining intensity using brightfield microscopy for each marker and thickness(n = 16). Immunostained sections of multiblock were scanned and the WSI was annotated by a pathologist. ROIs representing 100 cells or nuclei were selected (Fig. 2A). For each ROI, the staining intensity for 1+, 2+, or 3+ cells were scored by a pathologist (Fig. 2B). The ROIs annotated in the WSI were used to determine the stain absorbance range for all ROIs selected. The stain absorbance was used to determine a stain color distribution. Intensity cutoffs for the 4 bins that represent 0, 1+, 2+, or 3+ intensities (Fig. 2C) were determined by a pathologist on the stain color distribution map. Algorithmic analysis of each ROI yielded a mean stain absorbance (Fig. 2D) and percentage of pixels (Fig. 2E) assigned to each intensity bin. The pathologist score was then correlated to the algorithmically determined percentage of pixels (Tables 1, 2). This method was performed for Ki-67, BCL6, CD7, and CK immunostaining in sections cut at 2, 4, 6, or 8 μm. In addition, we compared the pathologist's scores manually using a brightfield microscope to the pathologist's visual score of the WSI. The pathologist's visual score of the WSI was highly correlated to the manual score by brightfield microscopy (data not shown; $r^2 = 0.99$).

Repeatability of Staining Intensity for Successive Cuts at a 6 μm Thickness

We first sought to determine whether serial sections of multiblock cut at 6 μm in thickness provided consistent results for staining intensity. We therefore immunostained 5 serial sections with antibody to Ki-67 nuclear antigen (Fig. 3). Tonsil tissue showed abundant immunoreactivity to Ki-67 within the nuclei of B cells in the germinal centers. The pathologist counted the number of 1+, 2+, and 3+ nuclei out of 100 nuclei in each ROI (n = 5 ROIs per tissue section) on the WSI (Fig. 3A). Algorithmic analysis was then performed to determine mean stain absorbance (Fig. 3B) and percentage of pixels (Fig. 3C) in the 0, 1+, 2+, and 3+ intensity bins. The results show a strong correlation between the pathologist's WSI score for the mean stain absorbance and the percentage of pixels per cutoff range for each intensity bin for all 5 tissue sections indicating a high level of repeatability (Table 1).

TABLE 1. Repeatability of Thickness Cut Measured by Ki-67 Number of Positive Nuclei and % of Pixels in Each Staining Intensity Bin in Tonsil Germinal Centers

Slide Number	Intensity Score	Mean % Nuclei	SD	Mean Percentage of Pixels	SD	r^2
1	0	ND	ND	5.94	4.32	0.890
	1+	3.00	3.32	15.85	8.57	
	2+	7.40	4.04	27.39	8.62	
	3+	89.60	6.58	50.82	20.94	
2	0	ND	ND	5.28	3.32	
	1+	3.40	1.95	14.54	7.58	
	2+	7.80	3.83	31.93	11.88	
	3+	88.80	5.12	48.25	20.87	
3	0	ND	ND	5.81	5.37	
	1+	4.20	2.59	15.20	8.33	
	2+	8.20	2.39	34.83	9.42	
	3+	87.60	3.51	44.16	20.77	
4	0	ND	ND	11.43	8.76	
	1+	3.40	1.52	15.88	9.70	
	2+	9.00	5.10	28.15	7.96	
	3+	87.60	6.11	44.54	25.42	
5	0	ND	ND	19.97	16.38	
	1+	3.40	2.41	16.24	11.03	
	2+	9.00	5.87	23.29	7.77	
	3+	87.60	7.92	40.50	27.10	

ND (no data): there is no data for 0 intensity in the visual scoring of whole-slide image.

Tissue Thickness Versus Staining Intensity in Sections Cut at Multiple Thicknesses

To discern whether there were measurable differences in staining intensity with different thickness of cuts, 4 sections of multiblock cut at either 2, 4, 6, or 8 μm were immunostained for Ki-67 (Figs. 2B, D, E), BCL6, CD7 or CK, (Fig. 4) and WSI were annotated for ROIs and analyzed using the algorithm. Ki-67 nuclear antigen expression in the nuclei of B cells in the germinal centers of tonsil tissue was scored by a pathologist for the number of nuclei with 1+, 2+, or 3+ intensity in a ROI (n = 10 ROIs per tissue section) of 100 nuclei using the WSI for tissues cut at different thickness (Fig. 2B andTable 2). As the thickness increases, more pixels are binned at the 3+ intensity and fewer are binned at the 1+ and 2+ intensity showing an increase in intensity with thickness. There was a statistically significant decrease in the number of 1+ intensity nuclei ($P < 0.025$) and 2+ intensity nuclei ($P < 0.05$) and a significant increase in the number of 3+ intensity nuclei ($P < 0.05$) for the 6 μm section compared with the 4 μm section.

The mean stain absorbance of the annotated ROIs (Fig. 2D) was calculated within each intensity bin and was consistent across thickness indicating that the algorithmic method was consistent with pathologist visual determination of the intensity range assigned to each bin based on the WSI stain color distribution. Supplemental Figure 1 shows the same results for BCL6, CD7, and CK (Figure, Supplemental Digital Content 1, http://links.lww.com/AIMM/A190). For the percentage of pixels within each bin (Fig. 2E), we see significant differences between thicknesses. There was a stepwise decrease in the percentage of pixels in the 0 and 1+ intensity bins and a significant increase in the percentage of pixels in the 2+ and 3+ intensity bins for the 6 μm cut compared with the 4 μm section ($P < 0.05$; Fig. 2E and Table 2).The results from the pathologist's visual scoring for Ki-67 and percentage of pixels were strongly correlated for 2, 4, 6, and 8 μm sections (Table 2).

BCL6 immunoreactivity was evaluated in the nuclei of B-cell rich regions in the germinal centers of tonsil tissue (Figs. 4A, B). BCL6 was scored using the WSI for sections at 2, 4, 6, and 8 μm thickness (n = 7 ROIs per tissue section). It is important to note that the 2 μm section had 0 intensity for BCL6 immunoreactivity and was therefore not evaluable for manual scoring. Tissue architecture was

TABLE 2. Visual Scoring and Algorithm Defined Percentage of Pixels in Each Staining Intensity Bin Results for Ki-67, BCL6, CD7, and CK for Tissues Cut at 4 Different Thicknesses

Antigen	"As-Cut" Tissue Thickness (μm)	Intensity Score	Mean % of Nuclei or Cells	SD	Mean Percentage of Pixels	SD	r^2
	2	0	ND	ND	22.20	11.22	0.907
		1+	16.31	9.53	33.09	15.22	
		2+	14.69	6.16	34.06	9.15	
		3+	69.00	7.77	10.65	10.70	
	4	0	ND	ND	26.44	10.46	0.929
		1+	11.43	5.93	25.34	4.31	
		2+	15.21	4.84	25.91	2.71	
Ki-67		3+	73.36	7.61	22.31	12.75	
	6	0	ND	ND	12.63	4.19	0.922
		1+	5.20	4.80	20.09	4.35	
		2+	11.50	4.48	28.17	4.63	
		3+	83.30	8.76	39.12	11.41	
	8	0	ND	ND	9.29	4.37	0.997
		1+	5.80	3.71	21.60	5.53	
		2+	11.70	5.09	33.73	6.53	
		3+	82.50	8.23	35.37	12.66	
	2	0	ND	ND	93.50	5.14	NA
		1+	0.00	0.00	5.84	4.36	
		2+	0.00	0.00	0.61	0.74	
		3+	0.00	0.00	0.04	0.06	
	4	0	ND	ND	76.74	10.39	0.981
		1+	65.50	19.78	16.02	4.73	
		2+	28.00	16.19	6.20	4.82	
BCL6		3+	6.50	4.12	1.05	1.34	
	6	0	ND	ND	49.52	13.37	0.959
		1+	38.80	16.12	24.96	2.39	
		2+	52.50	13.99	18.14	6.89	
		3+	8.70	6.15	7.39	7.92	
	8	0	ND	ND	33.43	9.71	0.970
		1+	27.30	17.22	23.00	4.63	
		2+	57.00	15.31	25.27	4.58	
		3+	15.70	5.36	18.30	10.98	
	2	0	ND	ND	59.89	12.56	0.904
		1+	35.00	16.58	25.81	6.60	
		2+	39.00	6.52	10.65	4.42	
		3+	26.00	17.82	3.65	2.08	
	4	0	ND	ND	37.74	1.74	0.884
		1+	35.00	10.00	26.91	1.24	
		2+	35.00	10.00	18.40	0.69	
CD7		3+	30.00	7.07	16.95	2.38	
	6	0	ND	ND	31.01	38.80	0.952
		1+	7.00	2.74	17.43	10.18	
		2+	19.00	8.22	18.43	10.52	
		3+	74.00	10.84	33.13	20.45	
	8	0	ND	ND	18.04	12.29	0.966
		1+	3.00	2.74	23.09	7.94	
		2+	16.00	2.24	24.60	3.83	
		3+	81.00	4.18	34.26	15.52	
	2	0	ND	ND	77.78	14.41	0.839
		1+	7.50	2.89	22.13	14.42	
		2+	82.50	5.00	0.09	0.03	
		3+	10.00	4.08	0.01	0.00	
	4	0	ND	ND	64.36	7.77	0.946
		1+	6.25	4.79	34.96	7.47	
		2+	72.50	16.58	0.56	0.27	
CK*		3+	21.25	13.15	0.12	0.09	
	6	0	ND	ND	51.21	10.35	0.895
		1+	0.00	0.00	47.26	9.51	
		2+	48.75	6.29	1.31	0.78	
		3+	51.25	6.29	0.22	0.13	
	8	0	ND	ND	32.26	8.76	0.898
		1+	0.00	0.00	63.65	5.90	
		2+	65.00	23.45	3.26	2.32	
		3+	35.00	23.45	0.83	0.77	

ND (no data): there is no data for 0 intensity in the visual scoring of WSI.

NA (not available): BCL6 scored at a 0 intensity by visual scoring of WSI. We could not correlate data to pixel intensity.

*For CK correlation for the 6 and 8 μm sections the 3+ and 2+ intensities were correlated to the percentage of pixels as the visual score for the 1+ intensity had no 1+ intensity cells.

CK indicates cytokeratin; WSI, whole-slide image.

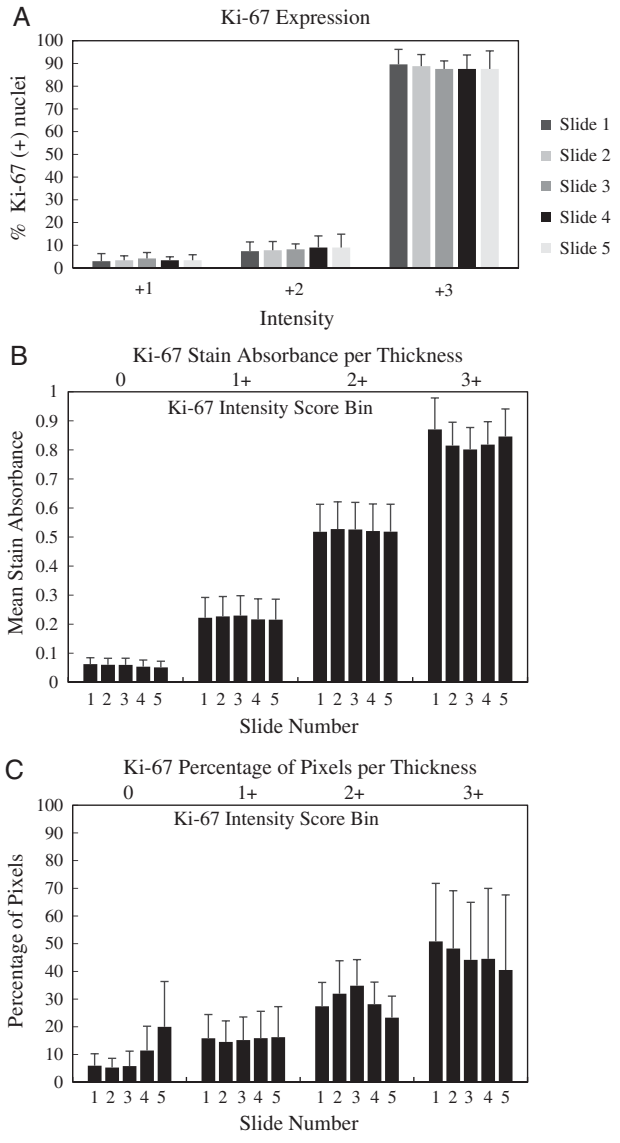

FIGURE 3. Determination of repeatability of staining intensity measurements in 5 serial sections of tonsil tissue cut at 6 μm. Five serial sections of multiblock formalin-fixed, paraffin-embedded tissues were cut at 6 μm of thickness and then immunostained with Ki-67 nuclear antigen. WSIs were then annotated and a stain color distribution map was derived from the data (Fig. 2C). The staining intensity cutoffs were then determined visually for the algorithmic analysis. The visual score for Ki-67 immunostaining at 1+, 2+, or 3+ intensity bin was determined for 5 regions of interest containing 100 B cells each in germinal centers of tonsil tissue (A). Data shown are mean values (± SD). The mean stain absorbance (B) and percentage of pixels (C) in each cutoff window were determined by algorithmic analysis. Visual scoring of WSI and algorithmic analysis showed a close correlation ($r^2 = 0.890$) in the consistency of 1+, 2+, and 3+ intensity ranges. WSI indicates whole-slide images.

faintly visible by inspection of the WSI which was annotated for ROIs that represent 100 nuclei (Figure, Supplemental Digital Content 2, http://links.lww.com/AIMM/A191) and the ROIs were used to calculate the percentage of pixels within each intensity bin (Fig. 4B). For BCL6, there was an increase in the number of nuclei with high (2+ and 3+) intensity with increased tissue thickness (Fig. 4A). There was a significant decrease in the number of 1+ nuclei and a significant increase in the number of 2+ nuclei in the 6 μm sections as compared with the 4 μm section ($P < 0.05$).

Surprisingly, variations in the percentage of pixels was measurable by algorithmic analysis in the 0 intensity bin. The algorithmic analysis showed the presence of pixels not discernable by manual scoring. There was an increased percentage of pixels with increasing thickness. Similar to Ki-67 results, there was a stepwise significant increase in the percentage of pixels in the 2+ and 3+ intensity bins with increasing tissue thickness ($P < 0.05$; Fig. 4B). The pathologist's score was strongly correlated with the percent of pixels within each intensity bin (Table 2). It is important to note that the algorithmic method could detect changes within the 2 μm cut section that was assigned a visual staining intensity of 0. This indicates that the algorithmic analysis method is a more sensitive tool for the evaluation of the percentage of pixels per unit area than the human eye.

CD7 immunoreactivity was evaluated in the membrane of T-cell rich MALT in tonsil tissue. WSI was annotated for ROIs (n = 3 ROIs per tissue section) representing 100 cells expressing CD7 at the 1+, 2+, or 3+ intensity scored using the WSI (Figure, Supplemental Digital Content 2, http://links.lww.com/AIMM/A191). By manual scoring, the number of CD7 positive cells with high staining intensity increased with thickness (Fig. 4C). There was a significant change in the number of 2+ and 3+ cells in the 6 μm section as compared with the 4 μm section ($P < 0.025$). The percentage of pixels trended toward decreased values in the 0 intensity bin with thickness (Fig. 4D). For CD7 the percentage of pixels increase significantly in the 2+ and 3+ intensity bins with increased thickness. The visual score correlated closely with the percentage of pixels for each WSI (Table 2).

CK expression was evaluated in the cytoplasm and membrane of hepatocytes in liver tissue cut at each thickness. The WSIs were annotated for ROIs (n = 4 ROIs per tissue section) that contained cells with 1+, 2+, and 3+ intensities (Figure, Supplemental Digital Content 2, http://links.lww.com/AIMM/A191) and the percentage of pixels in each intensity bin (Fig. 4F) was determined algorithmically. For the manual score (Fig. 4E), there were no detectable 1+ intensity positive cells visible in the 6 or 8 μm sections. Overall, CK expression showed a stepwise increase in the number of 3+ intensity cells with increased thickness. There was a significant increase in 2+ intensity positive cells in the 8 μm section as compared with the 6 μm section ($P < 0.05$). Similar to all markers used in this study, the percentage of pixels for CK decreased in the 0 intensity bin in a stepwise manner with increased thickness and increased in the 1+, 2+, and 3+ intensity bins with increased thickness. CK manual score and percentage of pixels were highly correlated (Table 2).

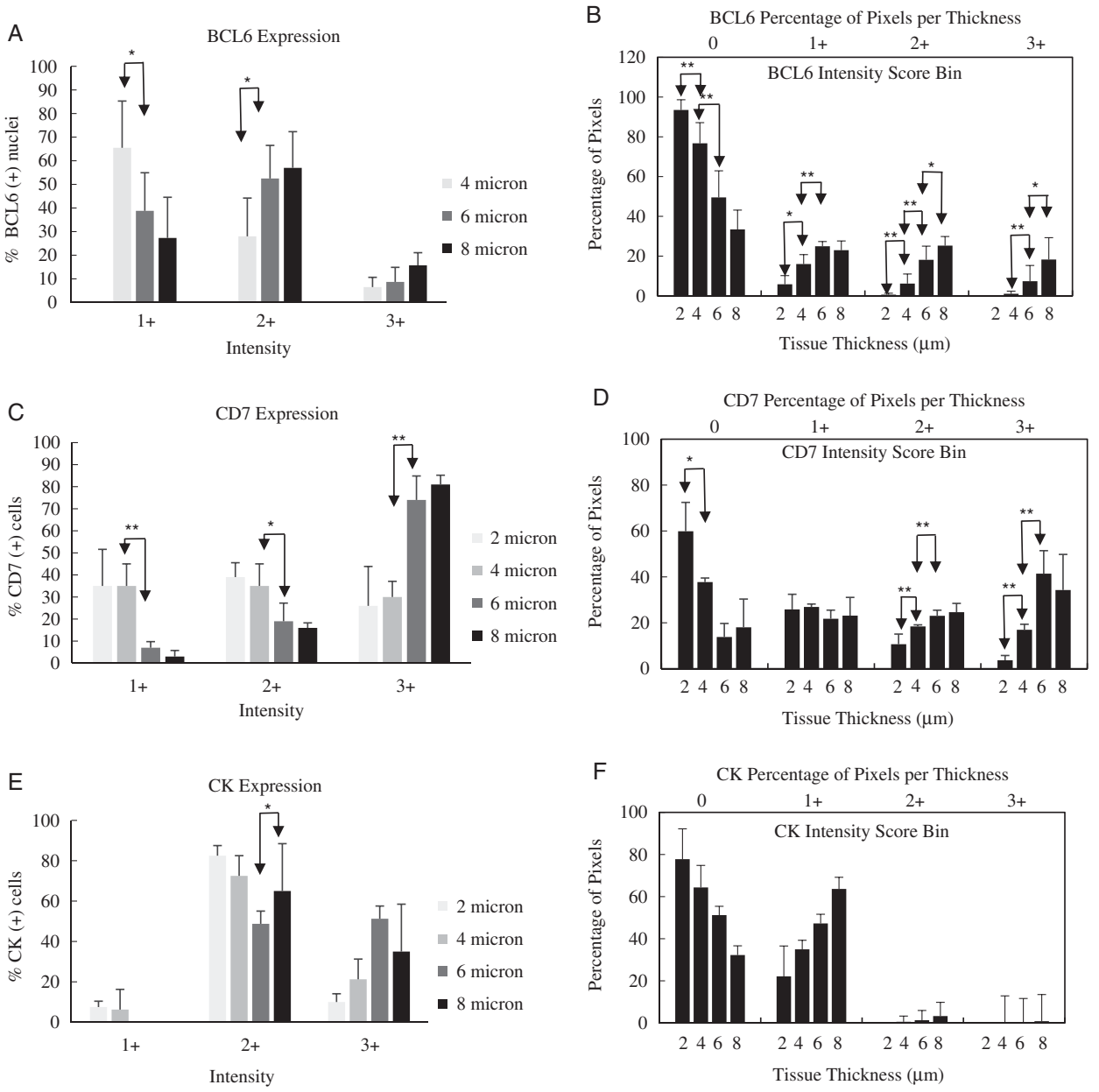

FIGURE 4. Determination of differences in staining intensity with tissue thickness for BCL6, CD7, and cytokeratin (CK). Sections of multiblock formalin-fixed, paraffin-embedded tissues were cut at 2, 4, 6, and 8 mm of thickness and then immunostained with antibodies to BCL6, CD7, and CK. Whole-slide images were annotated for ROIs and a manual score was determined for the number of cells within the ROIs that immunostained at a 1+, 2+, or 3+ intensity bin for BCL6 (A), CD7 (C), and CK (E). The percentage of pixels in 0, 1+, 2+, and 3+ intensity bins was determined by algorithmic analysis for BCL6 (B), CD7 (D), and CK (F). Data shown are mean values (± SD). *Statistically significant at the $P < 0.05$ level. **Statistically significant at the $P < 0.025$ level. ROI indicates region of interest.

A comparison of the pathologist's visual scores for the 3 intensities (1+, 2+, and 3+) to the percentage of pixels for the 4 intensity bins (0, 1+, 2+, and 3+) show variability in the trend to either increase or decrease with thickness. To account for variability between the pathologist score and the digital image analysis score or the comparison of 3 categories of pathologist visual score to 4 categories of algorithmically defined percentage of pixel's data we combined the low-intensity bins (0 and 1+) and the high-intensity bins (2+ and 3+). The comparison is a more simplified scoring system based on the scoring method that a 2+ to 3+ intensity in > 10% of cells would indicate positive expression for the corresponding antibody and a 0 to 1+ intensity would be scored as negative for antibody expression. For example, Ki-67 as scored by this traditional scoring method is technically positive

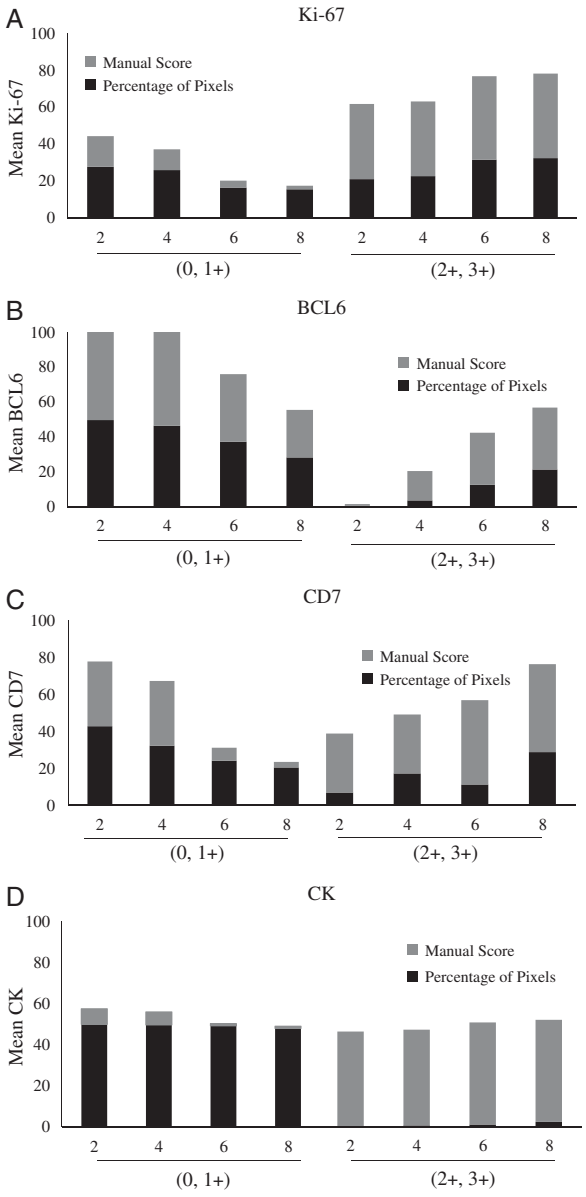

FIGURE 5. Staining intensity increases with tissue thickness for Ki-67 (A), BCL6 (B), CD7 (C), and CK (D) according to manual scoring and algorithmically defined percentage of pixels. Mean values for the percentage of pixels in the 0 and 1+ intensity bins were compared with the mean values for the 1+ intensity manual scores. Mean values for the percentage of pixels in the 2+ and 3+ intensity bins were compared with the mean values for the manual scores of 2+ and 3+ intensity cells or nuclei. The manual score for 1+ intensity cells or nuclei followed the same downward trend with increasing thickness as the percentage of pixels in the 0 and 1+ bins for all antibodies. Both the manual score for 2+ and 3+ intensity cells or nuclei and the percentage of pixels in the 2+ and 3+ intensities increased with increased tissue thickness. CK indicates cytokeratin.

(2+ to 3+ intensity in >10% of cells) in all 4 section thicknesses, whereas BCL6 would be scored as negative for the 2 μm cut tissue section. Figure 5 is a summary of the mean scores from manual and algorithmic scoring methods for Ki-67 (Fig. 5A), BCL6 (Fig. 5B), CD7 (Fig. 5C), and CK (Fig. 5D). Both manual and algorithmic defined staining intensity increased in a stepwise manner with increasing tissue thickness. If 0 and 1+ are defined as negative and 2+ and 3+ are defined as positive cells or nuclei, Figure 5 shows increased positivity with increasing thickness.

DISCUSSION

The quantification of the thickness by interferometry of microtomy cut FFPE tissues that have been stained by IHC has been previously performed in the interest of improving standardization of IHC marker testing. Barker et al[13] used interferometry to evaluate the properties of section thickness to guarantee the quality of FFPE MDA-MB-231 breast carcinoma control cell line sections specifically for IHC-based HER2 testing assay. They characterized HER2 IHC staining with the microtome settings of 3, 5, and 7 μm. They identified variations in the effective thickness of tissue sections that then affected the interpretation of HER2 IHC-based intensity of staining. Specifically, when sectioned at 3 and 5 μm, HER2 expression was 0 to 1+ in intensity, whereas in serial sections at 7 μm thickness 2 of 20 specimens increased their score to 2+, a borderline score for HER2 that would require reflexive molecular testing to verify HER2 status. Our study found that for 50 sections each of multiblock cut at 2, 4, 6, and 8 μm in thickness according to the microtome settings, there was minimal variation in thickness on the paraffin at the paraffin-slide interface (>1 μm) as determined by interferometry. Interestingly, our study identified a significant difference by manual scoring for IHC staining intensity and algorithmic defined percentage of pixels between 4 and 6 μm cut sections for 3 of the 4 markers tested. Similar to the findings of Barker and colleagues our data also suggest that a 2 μm difference in the thickness cut can affect staining intensity to the degree that a marker may be incorrectly categorized as positive or negative for immunoreactivity.

We then investigated whether subtle changes in immunohistochemical staining with changes in tissue thickness would be detectable to the pathologist and alternately using an algorithmic method to determine the percentage of pixels in the WSI. We interrogated the expression of nuclear (Ki-67 and BCL6), membranous (CD7), and cytoplasmic (CK) localized antibodies in multiple tissue types cut at different tissue thickness by annotating very small defined ROIs using a WSI and comparing manual scoring methods to our algorithmically defined percentage of pixels. Our data show that differences in staining intensity is quantifiable as a manual score of 1+, 2+, and 3+ intensity immunostained nuclei or cells in a defined ROI of 100 nuclei or cells expressing the protein of interest. The algorithmic method showed that the percentage of pixels staining within an intensity bin within that given area can be quantified and are highly correlated to the manual score.

Surprisingly, the algorithmic technique could detect a difference in the percentage of pixels between BCL6 immunostained 2 μm section of tonsil tissue that was indiscernible to the pathologist. This 2 μm section did not have a visible intensity >0 for BCL6 immunoreactivity by brightfield microscopy or by viewing of the WSI. The tissue appeared devoid of staining, although the underlying architecture of the tonsil tissue was faintly visible. Annotation of ROIs over the surface of the tonsil tissue, did show that the use of the digital image analysis could detect pixels in the WSI of this tissue. Table 2 indicates that the percentage of pixels for this section that fall into the 0 intensity bin is the majority at 93.5%, however, the remaining 6.5% of pixels are distributed between the 1+ intensity (5.84%), 2+ intensity (0.61%), and 3+ intensity (0.04%) bins. Thus, our digital image analysis was extremely sensitive to very subtle differences in the percentage of pixels that are not quantifiable visually. With further development, we believe that our algorithmic characterization technique could potentially be applied to IHC stain for WSI scoring applications. Our data does indicate that there is a positive impact of scoring standardization with the use of algorithms; however, to harvest such benefits, the most basic preanalytic factors need to be in place.

The ASCO and CAP have published Pathologists Guidelines for IHC testing of biomarkers for multiple cancers including breast, CNS, thyroid, colon, rectum, endometrium, hematologic, melanoma, and lung.[14–24] These templates are designed to assist pathologist in providing clinically useful and relevant information when reporting results from biomarker testing for IHC markers and also genetic mutation analysis.[25] Required elements of the templates include detailed information regarding IHC biomarker percentage of positive tumor cells, average intensity of staining (weak, moderate, and strong) and internal controls as well as preanalytic variables such as cold ischemia and fixation times, fixative type, adequacy of sample for evaluation, biomarker antibody clone, and image analysis assistance.[25] However, these templates do not include an indication of the thickness of tissue section cut for analysis. In addition, CAP guidelines for the handling of surgical pathology specimens also calls for routine tissues to be cut at 4 to 5 μm in thickness at microtomy.[5] Our results from this study indicate that tissue sections used for biomarker scoring be cut consistently at the same thickness of tissue on the slide with an error as small as 2 μm deleteriously effecting IHC interpretation of scoring by the pathologist. Further, it would be valuable for the histology laboratory to document the "as-cut" thickness for the pathologist in the standardized methods to draw attention to its importance. Staining intensity is indeed dependent on tissue thickness and sections should be cut within the middle of the range established by the assay manufacturer. In addition, our studies demonstrate the need for recommendations towards consistency of the thickness of the cut of specimens in the methodology by the International Ki-67 in Breast Cancer Working Group and other parties working towards harmonization of methodology for prognostic biomarkers.

ACKNOWLEDGMENT

We would like to thank Dr Lars Rudbeck of Agilent Technologies Inc. for his editorial contributions to this manuscript.

REFERENCES

1. Onitilo AA, Engel JM, Greenlee RT, et al. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res*. 2009;7:4–13.
2. Leong TY, Cooper K, Leong AS. Immunohistology–past, present, and future. *Adv Anat Pathol*. 2010;17:404–418.
3. Leong AS. Quantitative immunohistology: tissue section thickness, another glitch in the path to standardization. *Appl Immunohistochem Mol Morphol*. 2009;17:465–469.
4. Hammond ME, Hicks DG. American Society of Clinical Oncology/College of American Pathologists human epidermal growth factor receptor 2 testing clinical practice guideline upcoming modifications: proof that clinical practice guidelines are living documents. *Arch Pathol Lab Med*. 2015;139:970–971.
5. Lott R, Tunnicliffe J, Sheppard E, et al. Practical Guide to Specimen Handling in Surgical Pathology. *College of American Pathologists*. 2015;1–52.
6. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst*. 2011;103:1656–1664.
7. Abubakar M, Orr N, Daley F, et al. Prognostic value of automated KI67 scoring in breast cancer: a centralised evaluation of 8088 patients from 10 study groups. *Breast Cancer Res*. 2016;18:104–117.
8. Oh H, Eliassen AH, Wang M, et al. Expression of estrogen receptor, progesterone receptor, and Ki67 in normal breast tissue in relation to subsequent risk of breast cancer. *NPJ Breast Cancer*. 2016;2:1–13.
9. Chen M, Yao S, Cao Q, et al. The prognostic value of Ki67 in ovarian high-grade serous carcinoma: an 11-year cohort study of Chinese patients. *Oncotarget*. 2016. Doi: 10.18632/oncotarget.14112.
10. Melling N, Kowitz CM, Simon R, et al. High Ki67 expression is an independent good prognostic marker in colorectal cancer. *J Clin Pathol*. 2016;69:209–214.
11. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst*. 2013;105:1897–1906.
12. Anthony A, Colurso GJ, Bocan TM, et al. Interferometric analysis of intrasection and intersection thickness variability associated with cryostat microtomy. *Histochem J*. 1984;16:61–70.
13. Barker C, Ibrahim M, Miller K, et al. Nondestructive quality control of HER2 control cell line sections: the use of interferometry for measuring section thickness and implications for HER2 interpretation on breast tissue. *Appl Immunohistochem Mol Morphol*. 2009;17:536–542.
14. Fitzgibbons PL, Dillon DA, Alsabeh R, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. *Arch Pathol Lab Med*. 2014;138:595–601.
15. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134:e48–e72.
16. Bartley AN, Hamilton SR, Alsabeh R, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the colon and rectum. *Arch Pathol Lab Med*. 2014;138:166–170.
17. Cagle PT, Sholl LM, Lindeman NI, et al. Template for reporting results of biomarker testing of specimens from patients with non-small cell carcinoma of the lung. *Arch Pathol Lab Med*. 2014;138:171–174.
18. Hameed M, Corless C, George S, et al. Template for reporting results of biomarker testing of specimens from patients with gastrointestinal stromal tumors. *Arch Pathol Lab Med*. 2015;139:1271–1275.
19. Sholl LM, Andea A, Bridge JA, et al. Template for reporting results of biomarker testing of specimens from patients with melanoma. *Arch Pathol Lab Med*. 2016;140:355–357.

20. Brat DJ, Cagle PT, Dillon DA, et al. Template for reporting results of biomarker testing of specimens from patients with tumors of the central nervous system. *Arch Pathol Lab Med*. 2015;139:1087–1093.
21. Duncavage E, Advani RH, Agosti S, et al. Template for Reporting Results of Biomarker Testing of Specimens From Patients With Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma. *Arch Pathol Lab Med*. 2016;140:1228–1230.
22. Duncavage E, Advani RH, Agosti S, et al. Template for reporting results of biomarker testing of specimens from patients with diffuse large b-cell lymphoma, not otherwise specified. *Arch Pathol Lab Med*. 2016;140:1225–1227.
23. Kelley TW, Arber DA, Gibson C, et al. Template for reporting results of monitoring tests for patients with chronic myelogenous leukemia (BCR-ABL1). *Arch Pathol Lab Med*. 2015;140:672–674.
24. Bartley AN, Christ J, Fitzgibbons PL, et al. Template for reporting results of HER2 (ERBB2) biomarker testing of specimens from patients with adenocarcinoma of the stomach or esophagogastric junction. *Arch Pathol Lab Med*. 2015;139:618–620.
25. Fitzgibbons PL, Lazar AJ, Spencer S. Introducing new College of American Pathologists reporting templates for cancer biomarkers. *Arch Pathol Lab Med*. 2014;138:157–158.