Digital Image Analysis and Quantitative Bead Standards in Root Cause Analysis of Immunohistochemical Staining Variability: A Real-world Example

Rebecca Rojansky, MD, PhD,* Seshi R. Sompuram, PhD,†
Ellen Gomulia, MS HTL (ASCP) QIHC,* Yasodha Natkunam, MD, PhD,*
Megan L. Troxell, MD, PhD,* and Sebastian Fernandez-Pol, MD, PhD*

Abstract: Assessment of automated immunohistochemical staining platform performance is largely limited to the visual evaluation of individual slides by trained personnel. Quantitative assessment of stain intensity is not typically performed. Here we describe our experience with 2 quantitative strategies that were instrumental in root cause investigations performed to identify the sources of suboptimal staining quality (decreased stain intensity and increased variability). In addition, these tools were utilized as adjuncts in validation of a new immunohistochemical staining instrument. The novel methods utilized in the investigation include quantitative assessment of whole slide images (WSI) and commercially available quantitative calibrators. Over the course of ~13 months, these methods helped to identify and verify correction of 2 sources of suboptimal staining. One root cause of suboptimal staining was insufficient/variable power delivery from our building's electrical circuit. This led us to use uninterruptible power managers for all automated immunostainer instruments, which restored expected stain intensity and consistency. Later, we encountered one instrument that, despite passing all vendor quality control checks and not showing error alerts was suspected of yielding suboptimal stain quality. WSI analysis and quantitative calibrators provided a clear evidence that proved critical in confirming the pathologists' visual impressions. This led to the replacement of the instrument, which was then validated using a combination of standard validation metrics supplemented by WSI analysis and quantitative calibrators. These root cause analyses document 2 variables that are critical in producing optimal immunohistochemical stain results and also provide real-world examples of how the application of quantitative tools to measure automated immunohistochemical stain output can provide a greater objectivity when assessing immunohistochemical stain quality.

Key Words: autostainer, calibrator, image analysis, immunohistochemistry, monitoring, quality assessment, quantitative metrics

In the current paradigm of immunohistochemical stain validation and quality control, a trained pathologist evaluates the patient tissue, on slide control tissue, and/or external control tissue to assess whether the pattern and intensity of staining are appropriate. In addition, automated immunostainers are programmed to alert laboratory technicians when an instrument malfunction has occurred. Visual evaluation of control tissue and instrument alerts can readily detect gross failure of a protocol, but it is likely that more subtle abnormalities of a protocol remain undetected. These more subtle abnormalities may not impact the results of many stains used for diagnosis, but for dim stains or stains in which assessment near the limit of detection is clinically meaningful (eg, HER2, ER, PR, and PD-L1), a mild decrement in immunohistochemical stain intensity may result in inappropriate patient management. In the clinical immunohistochemistry laboratory, quantitative assessment of quality control material is not routinely performed. To our knowledge, quantitative assessment of the baseline within-run, between-run, and between-laboratory variability for automated stainers has only recently been published.[1] However, immunohistochemistry laboratories do not routinely monitor within-run and between-run variability in such a rigorous manner. Thus, determination of run acceptablity depends on the visual assessment of low (near limit of detection) or intermediate level control tissue. Without a quantitative measure of stain intensity, it may not be clear to pathologists what degree of variability seen in a laboratory is acceptable versus harbingers of reduced sensitivity. This variation is particularly problematic for

predictive and therapeutic markers. The use of Levey-Jennings charts and Westgard rules as in clinical laboratories has been explored,[2] but the infrastructure necessary to apply these approaches to the clinical immunohistochemistry laboratory are not yet widely available.

In this report, we describe root cause analyses that identified 2 sources of decreased stain intensity and increased variability that may be underappreciated and may not be caught by standard quality control methods. As part of our root cause analyses, we utilized 2 approaches to more precisely assess stain intensity: digital image analysis and quantitative calibrators. These tools proved essential to confirm the pathologist's visual impression of decreased stain intensity and allowed the accurate assessment of interventions that were put in place to address the source of suboptimal immunohistochemistry (IHC) stain results. In the first root cause analysis, we identified insufficient electrical power as the source of variably diminished stain intensity. In the second root cause analysis, digital image analysis and quantitative calibrators also provided objective evidence that one instrument was not functioning optimally. This was the case despite thorough review, repair, and calibration of the instrument by trained technicians and passing the vendor's quality control measures. These root cause analyses highlight 2 likely underappreciated sources of suboptimal immunohistochemical stain performance and provide further rationale for the adoption of more quantitative assessment of IHC stain results in clinical laboratories to more precisely monitor instrument performance.

MATERIALS AND METHODS

Immunohistochemistry

Immunohistochemical staining was performed on 5 automated stainers that were all of the same make and model. Immunohistochemical stains described in the study were performed in the Stanford Clinical Immunohistochemistry Laboratory, which is a Clinical Laboratory Improvement Amendments of 1988-licensed laboratory. Glass slides utilized throughout this study were Cardinal Health SuperFrost Plus slides (M6146-PLUS, Cardinal Health, Inc. Dublin, OH). The stains described in this study represent protocols that are validated and used for clinical purpose. The anaplastic lymphoma kinase (ALK) 5A4 clone (Abcam, Cambridge, MA, United States, catalog #ab17127) was used at a dilution of 1:25, with high pH antigen retrieval for 30 minutes. Control tissue used for the ALK stain studies was derived from 2 tissue blocks, both containing 3 tissues: ALK-positive inflammatory myofibroblastic tumor (positive control), ALK-positive lung adenocarcinoma (positive control), and a leiomyoma (negative control). Of note, only the ALK-positive inflammatory myofibroblastic tumor was used in the digital image analysis as described below. The PD-L1 E1L3N clone was obtained from Cell Signaling Technology (Danvers MA, United States, catalog # 13684S) and used at a dilution of 1:500 with high pH antigen retrieval for 30 minutes. Control tissue consisted of placenta and tonsil.

Several other stains were also evaluated in the course of troubleshooting and analysis, but for both visual (pathologist) evaluation and the whole slide images analysis described below, we focused largely on the ALK stain, while quantitative bead analysis focused on PD-L1.

Digital Image Analysis

Digital image analysis was performed on ALK-stained control tissue (ALK-positive inflammatory myofibroblastic tumor), derived from 2 control blocks described above. WSI were generated by scanning slides on a Philips IntelliSite Pathology Solution Ultra (Philips Medical Systems International BV, Best, NL) as per manufacturer instructions. Of note, images generated by the Philips scanners are not altered automatically or manually and thus, the differences in intensity of signal between WSI also reflect differences in stain intensity rather than changes in settings. After slide scanning, screenshots of control tissue at ×5 magnification (1536×590 pixel area) were exported from the image management system. Tissue regions of interest were selected to be the same or similar across all slides for each control block. Snapshots were segmented in QuPath as brightfield (H-DAB) using the integrated watershed cell detection that smooths the input image by log transform, removes background with a threshold filter, and then applies a watershed transform.[3] The same settings were applied for segmentation of all images (background radius = 10, median radius = 0, sigma = 3.0, minArea = 10, maxArea = 1000, threshold = 0.05, maxBackground = 2.0, and cell expansion = 3). The average of maximum pixel intensity per slide was calculated by summing the maximum pixel intensity per segmented cell and dividing by the total number of cells detected ($I_{avg} = \frac{\sum_1^n Imax_n}{n}$). Other measures of overall intensity were also assessed, but the average of maximum pixel intensity per slide most clearly revealed a decreased stain intensity and most closely correlated with the visual impression of the laboratory directors.

PD-L1 Intracellular Domain IHC Calibrator and Control

Quantitative calibrators and controls for the PD-L1 intracellular epitope were kindly provided by Boston Cell Standards (BCS). The BCS calibrators and controls incorporate a peptide spanning into most of the intracellular domain of PD-L1. The use of peptide epitopes as controls or calibrators in lieu of a native protein was previously described.[2,4] The PD-L1 intracellular domain peptide includes the epitopes of monoclonal antibodies (mAbs) SP142, SP263, E1L3N, ZR3, and 73-10. The peptide was purchased from CS Bio, Menlo Park, CA and is 93% pure based on mass spectroscopy analysis by the manufacturer. The remaining 7% is comprised of similar peptides that, due to less than 100% incorporation during synthesis, may randomly lack an individual amino acid. Most of these slightly truncated peptides still likely incorporate the relevant epitopes.
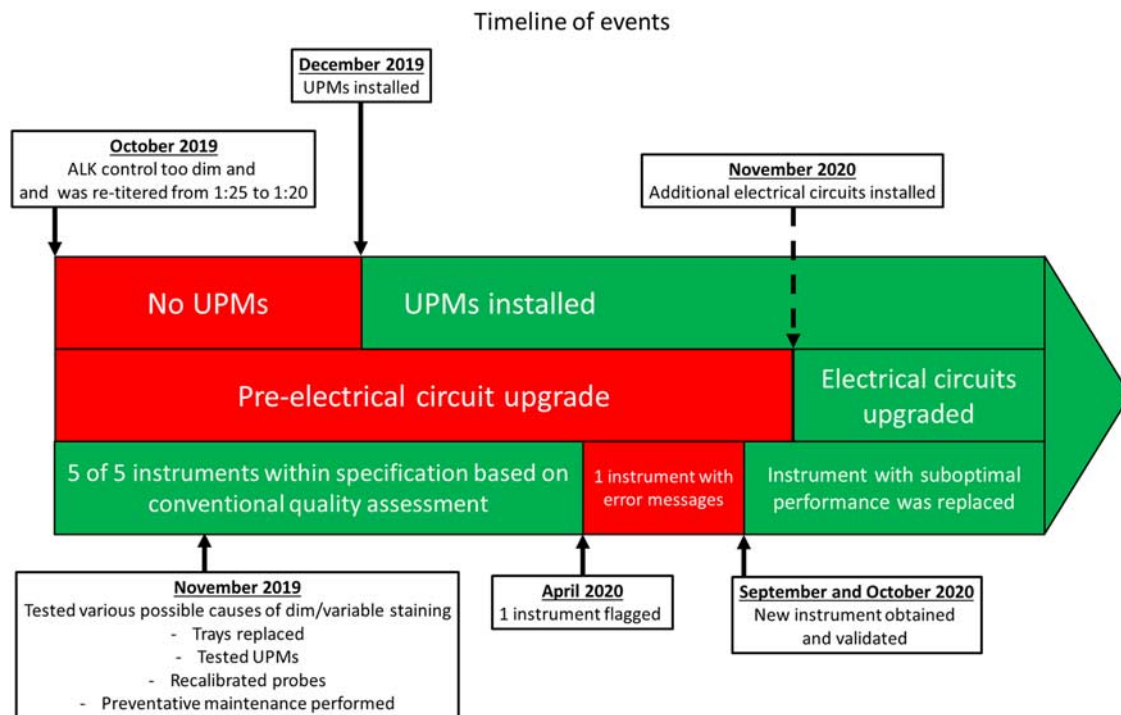
FIGURE 1. Timeline of events described in this report. Time is not to scale. ALK indicates anaplastic lymphoma kinase; UPM, uninterruptible power manager.

The intracellular domain BCS calibrators and controls were manufactured by covalently attaching a peptide to cell-sized (7-8-micron diameter) glass microbeads, as previously described.[5–7] For calibrators, 10 separate coupling reactions were performed at 10 different peptide concentrations, resulting in 10 different PD-L1 concentrations at regularly spaced concentration intervals. The peptide incorporates a fluorescein molecule to establish traceability of the measurement to NIST SRM 1934, as previously described.[8] For the intracellular domain PD-L1 BCS calibrators, those concentrations range from 34,000 to 2,200,000 molecules of PD-L1 peptide per microbead. These 10 different beads (corresponding to 10 different PD-L1 concentrations) are spotted onto a glass slide by the vendor to generate one calibrator slide. The PD-L1 calibrator slides were run on immunostainer instruments using the same PD-L1 immunohistochemical stain that is used for clinical care (E1L3N clone protocol described above). The stain intensity is proportional to the concentration of PD-L1 antigen and this is quantified using a digital image analysis workflow as previously described.[5] Briefly, the intensity of the stain is plotted versus the PD-L1 antigen concentration of the beads to generate a regression that shows how stain intensity varies by antigen concentration. A decrement in stain intensity (most notable with the higher concentration bead standards) compared with control instruments or conditions was used as an indicator that the staining process was suboptimal. If a calibrator slide was run and the intensity of stain for all points was indistinguishable from calibrators run on other instruments, this indicated that the stain performance was not compromised. Additional details of bead quantification have been previously described.[5,9] For control beads, the peptides are formalin-fixed in the presence of antigenically irrelevant proteins.[5–8,10,11] Two levels of controls (high and medium corresponding to

TABLE 1. Summary of Variables Tested During Troubleshooting and Their Impact on Stain Intensity and Consistency

Variables Evaluated/Tested	Impact
Antibody lot	No clear improvement
Antibody concentration	Mild increase in intensity with higher concentration
New slide trays (to control for tray warping)	No clear improvement
New covertiles	No clear improvement
Instrument calibration	No clear improvement
Preventative maintenance performed	No clear improvement
UPM	Improved intensity and decreased variability
Additional electrical circuit installed	Stain quality no longer impacted by UPM under test conditions

UPM indicates uninterruptible power manager.

different concentrations of PD-L1 antigen) are spotted onto glass slides. These control beads are different from the calibrator slide in that they can be placed anywhere on a slide and can serve as on slide control material in addition to or in lieu of on slide control tissue. For this study, control beads were spotted on slides that did not contain tissue in order to monitor within-run variability. As the calibrator beads take up an entire slide and are more expensive to make and analyze, control beads were used for analyses that require more slides (eg, within-run variability studies). Although control beads cannot be used to determine limit of detection, they can be used to detect suboptimal stain performance, which will manifest as the reduced intensity of staining. Both calibrator beads and control beads will only yield a signal when stained with antibodies that target epitopes within the peptide antigen that is coupled to the beads. Thus, a positive result indicates that the appropriate antibody was used and diminished signal or a negative result indicates that the staining process was compromised.

Statistical Analysis

Statistical analysis was performed in R (RStudio 4.0.3; R Foundation for Statistical Computing, Vienna, Austria). *P* values are derived from Student's paired *t* test. The variance is indicated as a percentage of the coefficient of variation (% CV), calculated as the standard deviation over the mean multiplied by 100. The threshold for significance was set to *P*-value < 0.001.

RESULTS

Root Cause Analysis 1: Identification of Electrical Power as a Source of Reduced Immunohistochemical Stain Intensity

Figure 1 provides a timeline of the relevant events mentioned in this study. In October 2019, routine review of our standard ALK on-slide controls revealed that the positive controls were dimmer than expected. Further testing with another lot of antibody also showed dimmer staining, thus, arguing against an antibody-specific cause (eg, reagent degradation). To determine whether the reduced intensity of staining was due to a factor related to the automated immunostainers, control tissue cut from the same block was tested on several instruments using fresh bulk reagents and freshly cut slides. In these tests, excessively dim and more varied intensity of staining was seen in a pattern that did not clearly suggest suboptimal performance of one particular instrument. Rather, the results of visual (1× and microscopic) assessment of ALK stains by the laboratory directors were most consistent with the conclusion that there was both within- and between-run variability for all stainers tested. As the root cause appeared to be a factor or factors intrinsic to the automated stainers, we tested the following variables as possible sources of dim and variable staining: tray warping, covertile age, component washing, instrument calibration, and preventative maintenance. Table 1 provides a list of variables tested and their impact on

stain quality. The introduction of uninterruptible power managers (UPMs), which are designed to deliver a consistent, surge-protected source of power, was found to restore the stain quality and consistency.

In order to confirm the pathologists' visual impression, we developed an objective tool to analyze DAB stain intensity from WSI. This tool quantifies the intensity of an immunohistochemical stain in a region of interest obtained from a WSI (see Materials and Methods). Figure 2 shows that the intensity as measured by digital image analysis was significantly lower in the absence of UPMs as compared with stain intensity with UPMs in place. The digital image analysis provides clear objective evidence that in the absence of UPMs stain intensity was lower and more variable (with UPM mean intensity 0.344, % coefficient of variation = 33.1%; without UPM mean intensity 0.149, % coefficient of variation = 39.6%).

Root Cause Analysis 2: Detection of Suboptimal Instrument Performance, Despite Conventional Quality Control

After UPMs were implemented, routine visual monitoring of immunohistochemical stain quality was as expected on all instruments (time period from December, 2019 to April 2020). However, in April 2020, one instrument (designated #108) showed an error message. The instrument was taken out of clinical use and the instrument vendor provided technical support to address the
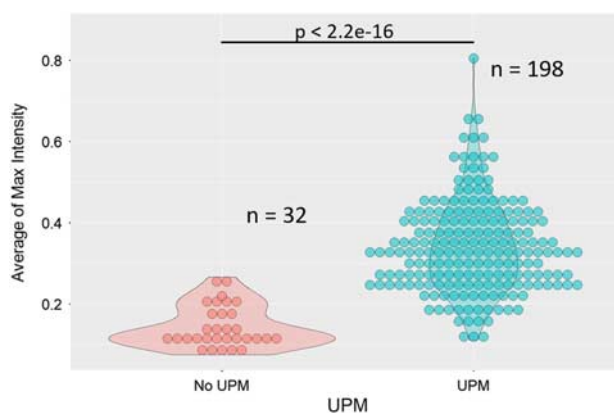

FIGURE 2. Impact of uninterruptible power managers (UPMs) on stain intensity and variability. Digital image analysis of ALK-stained control tissue supports the hypothesis that, before electrical circuit upgrades, in the absence of UPMs immunohistochemical stain intensity was reduced. The "Average of Max Intensity" represents the average of the maximum pixel intensity detected for each segmented cell. Each point represents the results of analyzing one tissue control (with one tissue control per slide). The no-UPM condition was performed as a dedicated test run (no patient samples). The UPM runs were cases that were run in the process of routine clinical care with on slide control material that was identical or similar to the control material used in the no-UPM condition. Details of the digital image analysis are provided in the Materials and Methods section. ALK indicates anaplastic lymphoma kinase.
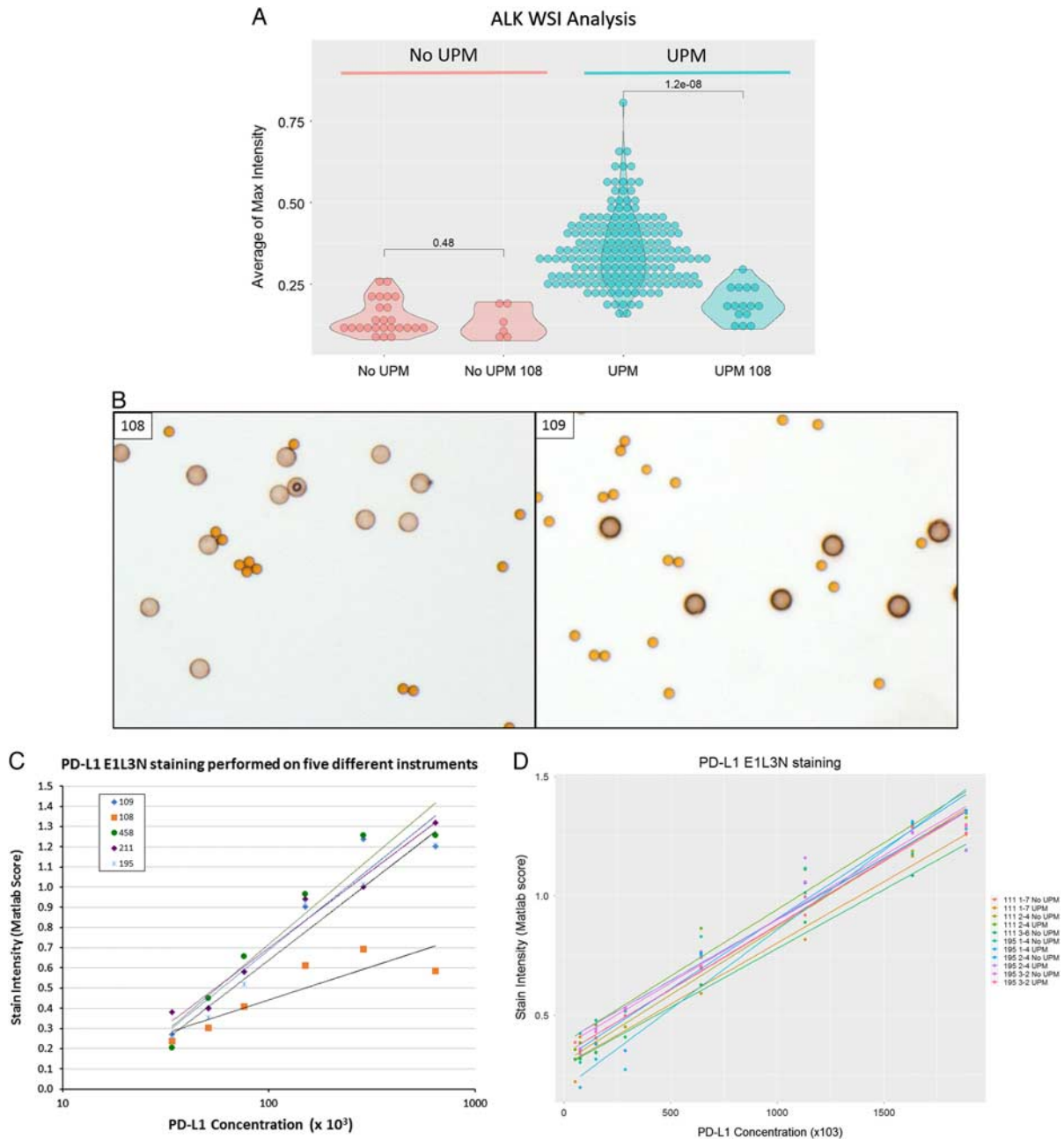
FIGURE 3. Whole slide images analysis and quantitative calibrators confirm persistent instrument malfunction. A, Whole slide image analysis shows that ALK stain intensity was lower for instrument #108 even when operating with a UPM. B and C, PD-L1 quantitative calibrators confirm that one instrument (designated #108) was yielding suboptimal staining (staining was performed with UPM in place). The E1L3N antibody was used as described in the Materials and Methods. B, Shows microscopic images are of calibrator beads stained on instrument 108 (defective instrument) and a comparator instrument 109. The small beads in both images are optical reference microbeads that are used for image quantification. The color of the reference microbeads is constant and not affected by staining. The larger circles are glass beads with a known number of PD-L1 peptides per bead. In these images, beads with 643,000 peptides per bead are shown. The E1L3N antibody–staining protocol shows dimmer staining when run on the defective instrument 108 (left image) when compared with a comparator instrument 109 (right image). The stain intensity seen with instrument 109 was visually indistinguishable from the other comparator instruments 195 and 458. C, Shows the analytic response curves of PD-L1 bead calibrators run on the 5 instruments in our laboratory. The legend indicates the instruments run, which are designated as 109, 108, 458, 211, and 195. The analytic response curve for instrument 108 clearly differs from those of the comparator instrument. D, The new instrument #111 shows E1L3N staining that is indistinguishable from comparator instrument #195. UPM did not impact staining intensity at the time of this run, which was performed after electrical circuit upgrades. The legend indicates instrument number (111 or 195) followed by tray position (eg, 1-7, 2-4, etc.) and then indicates the presence or absence of UPM. ALK indicates anaplastic lymphoma kinase.
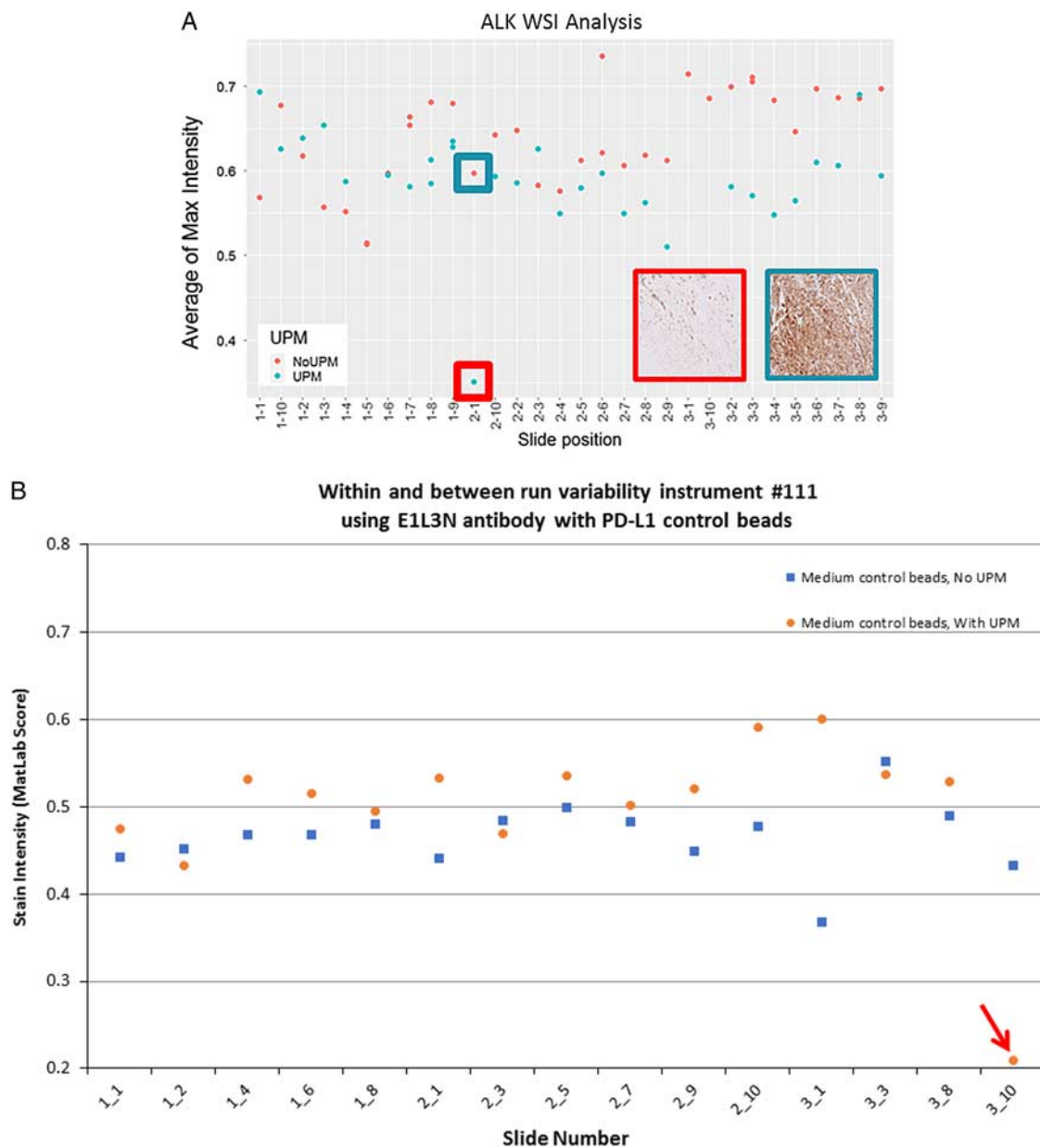
FIGURE 4. Within run variability seen with the whole slide image analysis method (A) and with bead control (E1L3N antibody with PD-L1 control beads) (B). All runs shown are after upgrade of the electrical circuits. In (A), whole slide image analysis of ALK-stained tissue run on instrument #195 shows that there is no consistent difference in stain intensity with or without UPM after the electrical circuit upgrade. Slide position 2-1 with UPM showed clearly a decreased staining by digital image analysis and visual evaluation confirmed lack of staining of the control tissue. B, Medium-level control beads were spotted onto glass slides as described in the Materials and Methods. Runs performed on instrument #111 with and without UPM in place are shown. The control beads showed reduced signal on the slide placed in position 3-10 (indicated by the red arrow). The instrument indicated an error for this slide. A repeat run did not result in a similar error (data not shown). Similar results were seen with high-level control beads that were spotted on the same slides (data not shown). ALK indicates anaplastic lymphoma kinase.

error. To further monitor instrument performance, we utilized standard pathologist evaluation (microscopic evaluation), the WSI digital analysis method, and quantitative calibrators.

Digital image analysis of ALK-stained control tissue suggested that instrument #108 continued to perform suboptimally. Figure 3A shows that even with a UPM installed and active, instrument #108 showed a lower intensity of staining compared with other instruments. In addition, PD-L1 calibrator beads showed that instrument #108 yielded dim stain intensity with our E1L3N PD-L1 stain protocol when compared with instruments of the

same make and model. As shown in Figure 3B, microscopic evaluation showed that PD-L1 calibrator beads stained on instrument #108 were dimmer than calibrator beads stained on other instruments of the same make and model. In Figure 3C, quantitative analysis of the calibrator beads shows that, as the peptide concentration of the calibrators increased, the intensity of signal generated with #108 was clearly lower than the intensity generated by the 4 other instruments used in the laboratory. In other words, the analytic response curves for the 4 properly functioning instruments were tightly concordant and instrument #108 was a clear outlier. Despite the vendor's thorough attempts to repair the instrument, suboptimal staining persisted and thus, it was agreed that an instrument replacement was needed to rectify stain quality.

Validation of New Instrument Using Digital Image Analysis and Quantitative Calibrators

As part of the validation of the new instrument, we utilized digital image analysis and quantitative calibrators to confirm that the new instrument (designated #111) yielded stains of similar intensity and consistency as the other 4 instruments. Figure 3D shows that stain intensity with quantitative calibrators run on the new instrument #111 was equivalent to the stain intensity run on a separate instrument (#195). In addition, the analytic response curves for calibrators run in different slide positions were essentially identical, thus, providing evidence that the various slide positions stained equivalently. Overall, standard instrument validation using pathologists' visual assessment, our digital image analysis method, and quantitative calibrators confirmed that the new instrument generated similar stain intensity and consistency as the other instruments in the laboratory.

Impact of Upgraded Circuitry on Stain Quality

In November 2020, new electrical circuits were installed in our laboratory as part of an upgrade to the building infrastructure to maintain sufficient power to instruments. We hypothesized that the presence of new electrical circuits may render the stain intensity less sensitive to the presence or absence of UPMs. Figure 4A shows results with our WSI analysis method, which is based on calculating the average of the maximum intensity pixels within all segmented cells. In the run shown, the average of maximum intensity without UPM across all 30 slides was 0.64, with a standard deviation of 0.058 and a %CV of 9.1%. With 1 outlier excluded (slide position 2-1 indicated in Fig. 4A) 29 ALK slides run with UPM in place resulted in a mean intensity of 0.59, with a standard deviation of 0.037 and a %CV of 6.2%. The difference between UPM and without UPM was no longer statistically significant (Student's t test P-value = 0.0006). Of note, the average of maximum intensity was higher in this experiment compared with earlier experiments due to the use of a new control block. Similar to the findings using ALK-stained tissue section, PD-L1 control microbeads spotted on slides stained with the E1L3N protocol also revealed that stain intensity was

identical with or without UPM (Fig. 4B). In addition, no slide position reproducibly showed a trend toward higher or lower stain intensity. These data demonstrate that in the presence of additional circuits, which likely provide appropriate electrical power more consistently, stain intensity is more consistent with or without UPMs. In addition, we did not identify a consistent difference in stain intensity across slide positions within either instrument.

DISCUSSION

The incorporation of quantitative metrics of stain intensity for instrument troubleshooting has not, to our knowledge, been previously reported. We describe 2 quantitative approaches used to enhance conventional visual assessment to identify root causes of suboptimal IHC stain quality: suboptimal power supply and suboptimal instrument performance. WSI analysis was possible because our clinical immunohistochemistry laboratory recently began scanning the large majority of slides using the Philips IntelliSite Pathology Solution Ultra. This allowed us to develop a semiautomated and more quantitative way of assessing stain intensity. A similar approach, using the Aperio ScanScope XT imaging system and a built-in DAB-specific–positive pixel count algorithm was recently utilized to study within-run, between-run, and between-laboratory variability by another group.[1] However, to our knowledge, there are no reports that describe the use of digital image analysis to troubleshoot immunohistochemical stain performance and aid in instrument validation. The second tools utilized in our study are quantitative calibrators and control beads that have a known quantity of antigen per bead.

Our experience with troubleshooting highlighted several weaknesses in the way that immunohistochemical stain quality and consistency is evaluated in clinical laboratories. One weakness is that the evaluation of on-slide control tissue from several runs over several days or longer time frames is generally not performed. Assessment of consistency across days and runs is dependent on the pathologists' visual impression of how intensely the control tissue typically stains, which can be subjective. Another weakness of the current paradigm for assessment of immunohistochemical stain quality and consistency is that it is at most semiquantitative and additionally depends on the stain being used for instrument evaluation. In troubleshooting instrument #108, we found that, with conventional visual assessment, certain stains, such as CD3 and CD10, were less sensitive than ALK for detecting diminished or variable stain intensity (data not shown). CD3 and CD10 tissue controls were insensitive to methodologic issues because the tonsil control tissue used to assess CD3- and CD10-containing antigen concentrations that are above the analytic response curve for those assays in our laboratory (and likely in most laboratories).[12] Our experience with conventional visual assessment confirmed that there is a significant potential for suboptimal instrument performance to be masked by using a tissue and antibody combination that result in a stain intensity outside the linear range.

TABLE 2. Comparison of Different Methods of Assessing Immunohistochemical Stain Results

	Visual Assessment of Control Tissue (Current Method Used by All Laboratories)	Digital Image Analysis of Control Tissue	Control Microbeads (High and Low Levels)	Calibrator Microbeads
Summary of workflow	Identify appropriate cases from archives	Identify appropriate cases from archives	Purchase microbead controls	Purchase calibrator slides
	Locate and pull blocks	Locate and pull blocks	Manually pipette controls instead of mounting tissue controls	Process calibrators on immunostainer
	Block cutting	Block cutting	Process slides with patient samples on immunostainer	Evaluate calibrators visually or with digital image analysis (currently performed by the vendor)
	Stain slides	Stain slides	Stained slides reviewed by pathologist	
	Stained slides reviewed by pathologist	Whole slide images		
		Digital image analysis (quantification of stain intensity)		
Analyte concentration known?	No	No	Yes	Yes
Current uses	Detect gross assay problems or instrument failure	None	Detect gross assay problems or instrument failure)	Used to establish limit of detection
Potential future uses	Same as above	Fine discrimination of partial IHC assay problems	Routine demonstration that the correct IHC stain was performed and was performed with satisfactory results (expected intensity)	Determine LOD for ensuring interinstitutional alignment of assay performance
		Levey-Jennings analysis	With added programs for image analysis and routine quantitation, global comparisons within runs, between runs, and between instruments to identify subtle changes in IHC stain performance that may be go unnoticed by visual assessment	
Degree of quantitation of intensity	Limited, typical scale is to 0 to 3+	Precise but relative quantification	Precise but relative quantification	Intensity quantification used for calculating LOD and dynamic range
Sensitivity to detect subtle decrements and variability in stain intensity	Lowest	Moderate	Low if visually assessed	Highest
			Moderate if with digital image analysis	
Assess limit of detection?	No	No	No	Yes
Currently available for routine use	Yes	No (requires custom workflow)	Dependent on availability from vendor	Dependent on availability from vendor

IHC indicates immunohistochemistry; LOD, limit of detection.

A third weakness is that, until recently, there has not been an assessment of the expected consistency of immunohistochemical staining within runs, between runs, and between laboratories.[1] For this reason, it was initially unclear whether the variability in stain intensity seen between slides containing the same control tissue was normal or it was due to a true problem with the staining process. In our root cause analysis, WSI analysis, quantitative calibrators, and quantitation of control beads were invaluable in confirming that stain intensity was more variable under certain situations (absence of UPM, insufficient power, and suboptimal instrument). In the future, quantitative tools, such as WSI analysis and/or quantitative calibrators, could be used routinely to establish an expected range for these stains in a given laboratory and to monitor stains over time, even on standard on slide controls. These methods would be allowed for the implementation of quality control tools such as Levey-Jennings charts to identify outliers and trends that may merit further investigation.

On the basis of our experiences described in this study, we hypothesize that both software tools to facilitate assessment of IHC stain intensity on WSI and tools such as quantitative calibrators will prove useful for ongoing monitoring of IHC stain performance. A summary of the methods used here compared with the current standard visual assessment is shown in Table 2. We think that these tools are complementary and will likely both be useful in the future. The WSI approach takes advantage of tissue controls that are already commonly produced. The method can be used to analyze on-slide control tissue and control tissue on separate slides. In the future, it may be possible to monitor on-slide control tissue globally and in near real-time to rapidly detect changes in stain performance, similar to what can be done in the clinical chemistry laboratory with quality control material and monitoring of moving averages. One important point regarding the WSI approach is that expression and antigenicity of the target can vary from control block to block, which will limit its ability to compare stain performance between different laboratories. Although the absolute levels of intensity may not be comparable from block to block, our experience shows that the digital evaluation of stain intensity for a particular block can improve a laboratory's ability to detect subtle decrements in stain intensity that warrant further investigation. Challenges to implementation of the WSI approach include lack of readily available user-friendly software to routinely perform this analysis. Quantitative calibrators have the added advantage of allowing assessment of the limit of detection, which is critical for predictive markers and/or therapeutic targets, such as ER, PR, and HER2, in which very low levels of protein expression are clinically

relevant. Furthermore, quantitative calibrators do not suffer from some of the problems inherent in control tissue, such as limited supply and potentially high variability in expression or antigenicity from block to block and thus, could potentially be used to harmonize assay results between laboratories. Both methods have the potential to measurably improve IHC stain performance and identify errors earlier than the current paradigm in clinical immunodiagnostic laboratories.

In conclusion, we report the development and application of a quantitative WSI method to assess immunohistochemistry variability, and we have illustrated how use of such a method, together with quantitative bead calibrators and control beads, can aid in identification, testing, resolution, and monitoring of stain variability in immunohistochemistry laboratories.

ACKNOWLEDGMENTS

REFERENCES

1. Chlipala EA, Bendzinski CM, Dorner C, et al. An image analysis solution for quantification and determination of immunohistochemistry staining reproducibility. *Appl Immunohistochem Mol Morphol.* 2020;28:428–436.
2. Vani K, Sompuram SR, Naber SP, et al. Levey-Jennings analysis uncovers unsuspected causes of immunohistochemistry stain variability. *Appl Immunohistochem Mol Morphol.* 2016;24:688–694.
3. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep.* 2017;7: 16878.
4. Bogen SA. A root cause analysis into the high error rate in clinical immunohistochemistry. *Appl Immunohistochem Mol Morphol.* 2019; 27:329–338.
5. Sompuram SR, Vani K, Tracey B, et al. Standardizing immunohistochemistry: a new reference control for detecting staining problems. *J Histochem Cytochem Off J Histochem Soc.* 2015;63:681–690.
6. Sompuram SR, Kodela V, Ramanathan H, et al. Synthetic peptides identified from phage-displayed combinatorial libraries as immunodiagnostic assay surrogate quality-control targets. *Clin Chem.* 2002; 48:410–420.
7. Sompuram SR, Vani K, Hafer LJ, et al. Antibodies immunoreactive with formalin-fixed tissue antigens recognize linear protein epitopes. *Am J Clin Pathol.* 2006;125:82–90.
8. Torlakovic EE, Sompuram SR, Vani K, et al. Development and validation of measurement traceability for in situ immunoassays. *Clin Chem.* 2021;67:763–771.
9. Sompuram SR, Vani K, Schaedle AK, et al. Quantitative assessment of immunohistochemistry laboratory performance by measuring analytic response curves and limits of detection. *Arch Pathol Lab Med.* 2018;142:851–862.
10. Sompuram SR, Vani K, Bogen SA. A molecular model of antigen retrieval using a peptide array. *Am J Clin Pathol.* 2006;125:91–98.
11. Vani K, Sompuram SR, Schaedle AK, et al. Analytic response curves of clinical breast cancer IHC tests. *J Histochem Cytochem Off J Histochem Soc.* 2017;65:273–283.
12. Vani K, Sompuram SR, Schaedle AK, et al. The importance of epitope density in selecting a sensitive positive IHC control. *J Histochem Cytochem Off J Histochem Soc.* 2017;65:463–477.