Surgical and Experimental Pathology

RESEARCH Open Access

Computer-assisted tumor grading, validation of PD-L1 scoring, and quantification of CD8-positive immune cell density in urothelial carcinoma, a visual guide for pathologists using QuPath

Aline Rodrigues[1,2], Cleto Nogueira[2,3], Laura Cardoso Marinho[2], Guilherme Velozo[2], Juliana Sousa[2], Paulo Goberlanio Silva[1] and Fabio Tavora[2,3*]

Abstract

Background: Advances in digital imaging in pathology and the new capacity to scan high-quality images have change the way to practice and research in surgical pathology. QuPath is an open-source pathology software that offers a reproducible way to analyze quantified variables. We aimed to present the functionality of biomarker scoring using QuPath and provide a guide for the validation of pathologic grading using a series of cases of urothelial carcinomas.

Methods: Tissue microarrays of urothelial carcinomas were constructed and scanned. The images stained with HE, CD8 and PD-L1 immunohistochemistry were imported into QuPath and dearrayed. Training images were used to build a grade classifier and applied to all cases. Quantification of CD8 and PD-L1 was undertaken for each core using cytoplasmic and membrane color segmentation and output measurement and compared with pathologists semi-quantitative assessments.

Results: There was a good correlation between tumor grade by the pathologist and by QuPath software (Kappa agreement 0.73). For low-grade carcinomas (by the report and pathologist), the concordance was not as high. Of the 32 low-grade tumors, 22 were correctly classified as low-grade, but 11 (34%) were diagnosed as high-grade, with the high-grade to the low-grade ratio in these misclassified cases ranging from 0.41 to 0.58. The median ratio for bona fide high-grade carcinomas was 0.59. Some of the reasons the authors list as potential mimickers for high-grade cases are fulguration artifact, nuclear hyperchromasia, folded tissues, and inconsistency in staining. The correlation analysis between the software and the pathologist showed that the CD8 marker showed a moderate ($r = 0.595$) and statistically significant ($p < 0.001$) correlation. The internal consistency of this parameter showed an index of 0.470. The correlation analysis between the software and the pathologist showed that the PDL1 marker showed a robust ($r = 0.834$) and significant ($p < 0.001$) correlation. The internal consistency of this parameter showed a CCI of 0.851.

*Correspondence: ftavora@gmail.com
[2] Argos Laboratory, Av Santos Dumont 5753, #1607, Fortaleza CE, 60175047, Brazil
Full list of author information is available at the end of the article

Conclusions: We were able to demonstrate the utility of QuPath in identifying and scoring tumor cells and IHC quantification of two biomarkers. The protocol we present uses a free open-source platform to help researchers deal with imaging and data processing in the surgical pathology field.

Keywords: Digital pathology, QuPath, CD8, Tumor infiltrating lymphocytes, Programmed death-ligand 1 immunotherapy, Biomarker, Bladder cancer

Introduction

Since the 1970s and, more importantly, in the past 20 years, immunohistochemistry (IHC) has been proven to be an effective, well-established, and widely accepted method in the Surgical Pathology laboratory to aid in the conventional histochemical stains of tissue sections. Diagnosis of disease, particularly cancer, is based on the examination of cells by light microscopy and on detecting specific molecules, such as proteins and nucleic acids, which are routinely identified in biopsy tissues by antibody-binding or nucleic hybridization technologies. The techniques have improved, and now, hundreds of antibodies, probes, and reaction protocols are performed worldwide to identify specific cell types, subclassify tumors, provide information on cancer-specific drivers, estimate aggressiveness of neoplasms, predict response to therapy, identify infectious agents, and many other applications (Cimino-Mathews 2021; McGinnis et al. 2021; Oh and Mahalingam 2019; Ross et al. 2020; Shah et al. 2013; Shelton et al. 2017). Guidelines for the standardization and analytic validation of IHC have been established by the College of American Pathologists, with regular updates (Hardy et al. 2013; Satturwar et al. 2019).

According to specific tumor types and antibodies used, pathologists usually score findings in a semiquantitative fashion, based on marker distribution, percent of positive cells, and/or stain intensity. IHC slides are predominantly stained with DAB and are counterstained with hematoxylin. Examples of successful surgical pathology histories include assessing HER2 positivity in breast cancer, ki67 quantification in various neoplasms, and, more recently, the use of PD-L1 scoring (in tumor and immune cells) to predict response to immunotherapy in a plethora of cancer types (Cimino-Mathews 2021; Guo et al. 2020; McGinnis et al. 2021). In both research and the clinical scenario, the H-score and Allred scoring systems are decades-old and still used to assess the percentage and intensity of cell staining (Harvey et al. 1999; McCarty Jr. et al. 1986).

Several research studies have relied on tissue microarrays (TMA) combined with IHC, to allow high throughput analysis of multiple tumor samples (Ilyas et al. 2013; Loughrey et al. 2018; Gray et al. 2017; Gurgel et al. 2020; Ozbudak et al. 2009; Morais et al. 2019). With digital pathology and the capacity to scan whole slides with high-quality of images, new software and platforms have been created to help researches and pathologists quantify and classify tissues and tumors. QuPath is an open-source pathology analysis software that offers a reproducible way to analyze and provide quantified variables and permits training and subsequent classification of cells using automated digital algorithms (Bankhead et al. 2017; Humphries et al. 2021).

In this study, we aimed to assess the (1) feasibility of scanning and analyzing tissue microarrays for bladder cancer, (2) show the functionality of biomarker scoring using QuPath, and (3) provide a step-by-step guide for the validation of pathologic grading using a series of cases of urothelial carcinoma stained with HE, CD8 and PD-L1.

Methods

Patients and tumor samples

The cohort consisted of 1 hundred and 40 bladder cancer patients with samples originating from transurethral resections and radical cystectomies from a reference laboratory in genitourinary pathology. Demographic data available and collected included sex, age, smoking status and overall survival. Individuals with neoadjuvant, prior BCG therapy or radiation therapy were excluded. Tumors obtained from formalin-fixed, paraffin-embedded (FFPE) specimens were retrospectively collected from the files of a reference laboratory in Northeastern Brazil after receiving Institutional Review Board approval. Tissue sections were retrieved from dominant tumor foci. Tumor samples were classified as low or high-grade according to the grading system accepted by the WHO in 2016.

Tissue microarray (TMA) construction

Two tissue microarrays (TMAs) were constructed using a manual tissue arrayer as previously described (Gurgel et al. 2020). For each case, a 2 mm core of a representative area was selected from the hematoxylin-eosin (HE)-stained sections from bladder specimens by an experienced genitourinary pathologist. Tissue cores that contained < 100 tumor cells were removed from the analysis.

Immunohistochemical analyses

IHC was performed in a private laboratory that holds the Brazilian Society of Pathology quality accreditation program seal. PD-L1 and CD8 were chosen to prove the versatility and potential use of QuPath in a specific T cell marker (CD8) and PD-L1, which is the most vital marker to assess response to immunotherapy. These biomarkers were chosen as both are well characterized, and antibodies are in routine clinical use in most pathology laboratories. In addition, these markers show very different patterns. CD8 marks specifically a subclass of T-cells in a uniform manner, while PD-L1 stains tumors and immune cells in membranous and/or cytoplasmic ways with variable intensity.

Sequential sections were cut from each TMA block and mounted on charged microscopy glass slides for immunohistochemical staining. The PD-L1 IHC analysis (22C3 pharmDx, Agilent, clone 22c3) was performed using PT-Link (Dako PT100), followed by target recovery with EnVision™ FLEX pH6.0 buffer, using the Agilent Technologies®, USA visualization system in Autostainer Link 48® equipment. PD-L1 positivity was assessed according to the current protocol.

Tumor T-cell immune infiltrate characterization was performed by evaluating CD8 (DAKO, clone C8/144b) by immunohistochemistry in all TMAs. The density of CD8+ T cells was evaluated as the overall percentage of the area within the borders of the tumors covered by positive immune cells. This method is based on the proposal for a standardized method from the International Immuno-Oncology Biomarkers Working group (Hendry et al. 2017).

Morphometric and computer-assisted image analyses

TMA stained slides (HE, CD8, and PD-L1) were scanned at 40x magnification (Motic EasyScan Infinity 100 scanner, Vancouver, CA) and analyzed by two pathologists independently using a semiquantitative score. Both pathologists also scored the tumors as low and high-grade according to the current World Health Organization Classification.

The files were loaded onto a project in QuPath software (QuPath source code, documentation, links to the software download are available at https://qupath.github.io). QuPath's segmentation feature is able to detect thousands of cells, identify them as objects in a hierarchical manner below the annotation, TMA cores, or cases, and measure cell morphology and biomarker expression at the same time (Bankhead et al. 2017). The images were disarrayed, and all cores separated in a continuous fashion from 1 to 140. The *Cell Detection* command was used to identify all cells in all cores based on the optical density of nuclear hematoxylin staining. Next, using selected measurements of intensity and morphology of all cells, applying a two-way *Random Trees Classifier* to train QuPath interactively to distinguish tumor cells from the stroma and low-grade predominant from high-grade predominant tumor cores. This required drawing tumor cells by an experienced pathologist, and QuPath was able to provide immediate feedback across all tumors. Once the classification was adequate for training images,, the authors expanded the *Classifier* to identify all cells in all cores and estimate % of high-grade and low-grade carcinomas to compare o the original and confirmed pathologist classification. If the output from QuPath showed that high-grade cells were present in more than 10% of the tumor cells in each core, on a Excel spreadsheet, the tumor was considered high-grade by software analyses. We randomly selected 7 high-grade and 3 low-grade carcinoma cases as a training set for the algorithm, annotated by one of the authors, to identify low-grade, high-grade tumor cells and stroma. The hypothesis is to evaluate if using only 7% of the cores (10/140), QUPath would be able to classify the remainder of the tumors in the TMA cores accurately.

For CD8 density measurements, a *Simple Tissue Detection* was performed to assess tissue area from each TMA core, followed by CD8 cell counts with the *Positive Cell Detection* command within the software (parameters for each command were fine-tuned by one of the authors). These measurements allowed the output of CD8 positive cells / mm2 of tissue. The corresponding script was then applied to all TMA cores, and data were exported to statistical analyses software (JMP, 2021 SAS Institute Inc). This evaluation of a CD marker in the TMA core has been previously described (Loughrey et al. 2018). The output for this variable was the number of CD8+ cells per mm2, and it was compared with the analyses by a pathologist who blindly assessed each one of the cores and assessed a percentage semiquantitative score from 0 to 100% (overall area involved by infiltrating or peritumoral CD8+ lymphocytes).

For PD-L1 scoring, analyses were run based on the trained algorithm described above, and quantification of cytoplasmic positive (mean DAB staining) only in cells detected as tumors cells, and not stroma and other cells. QuPath output was able to to identify numbero of PDL1 positive cells/mm2 and, in addition, the H-score method was applied by the software, based on extent and intensity of cytoplasmic staining (1-3), multiplied by the percentage of cells positive, with a potential score ranging from 0 to 300 (Bankhead et al. 2017). A pathologist score PD-L1 on each core using the established Tumor Proportion Score (TPS) method. It was defined as the percentage of viable tumour cells showing partial or complete membrane PD-L1 staining at any intensity. Comparison

between pathologists' analyses and the computer-assisted measurements were assessed by agreement scores.

Statistical analysis

CD8 and PDL1 data were evaluated using Spearman's correlation and internal consistency analysis by calculating the interclass correlation coefficient. Additionally, the kappa coefficient was calculated for the cut-off points (High vs Low). Cohen's kappa ranged from −1.0 (total discordance) to +1.0 (total concordance). Values between 0.6 to 0.8 are related to adequate concordance and 0.8 to 1.0 are related to strongly concordance. Values lower than 0.6 are considered inadequate concordance. In all instances, a confidence level of 95% was adopted (SPSS® v20.0 for Windows).

Results

Clinicopathological data

The cohort consisted of 140 patients, but 9 were immediately excluded in the grade analyses due to a lack of tissue cores in the TMA histologic sections stained with hematoxylin and eosin (Fig. 1, 131 total, 105/76% male) with the mean age 74.2 years (57% over 75 years of age). Sections for immunohistochemistry did not have drop-out cores, and all 140 cores were used for analyses (140 cores total, mean age 71.4 years, 77% male). Per the original pathology reports, there were 99 high-grade urothelial carcinomas (75,5%) and 32 low-grade, which were confirmed by blind evaluation of each core by an experienced genitourinary pathologist. Per pathology reports, 83 cases (59%) were invasive at least into lamina propria (all high-grade lesions), while 56 were non-invasive – this data was not confirmed on tissue cores due to the impossibility of assessing invasion in TMAs. Patients with low-grade tumors were younger (mean 68,6 years) compared to ones with high-grade tumors (mean 76, $p = 0{,}0062$). Patients with invasive tumors also were older (mean 76,1 years) than the ones with non-invasive (mean 71 years, $p = 0{,}036$).

Cell detection

After dearraying the TMA and running a script to enumerate the tissue cores from 1 to 140, QuPath was able to treat each core as a separate unit in a hierarchical manner for output of data and measurements. One of the main basic functions of QUPath (and any surgical pathology analyses software for that matter) is to identify cells and calculate individual features. We run the *Cell Detection* command using hematoxylin optical density as the target for the nuclei, with 0,5uM pixel size and 3uM of cell expansion. Figure 2 shows one of the cores (#61) as an example. For this core, there were 30,204 cellular detections. More importantly, for each cell, the program can provide all measurements and export them to a spreadsheet. If cell delineation is acceptable, *Cell Detection* can be run for all tissue cores. The median number of detected cells for all cores was 25,921 cells (range 3197-64,144). A total of 3,283,282 cells were identified in all 131 cores. The authors used a Core i7 PC computer running Windows 10, with 16GB of RAM memory and a SSD hard disk of 512GB. For each TMA comprising 7 cores, *Cell Detection* run for about 3 min.

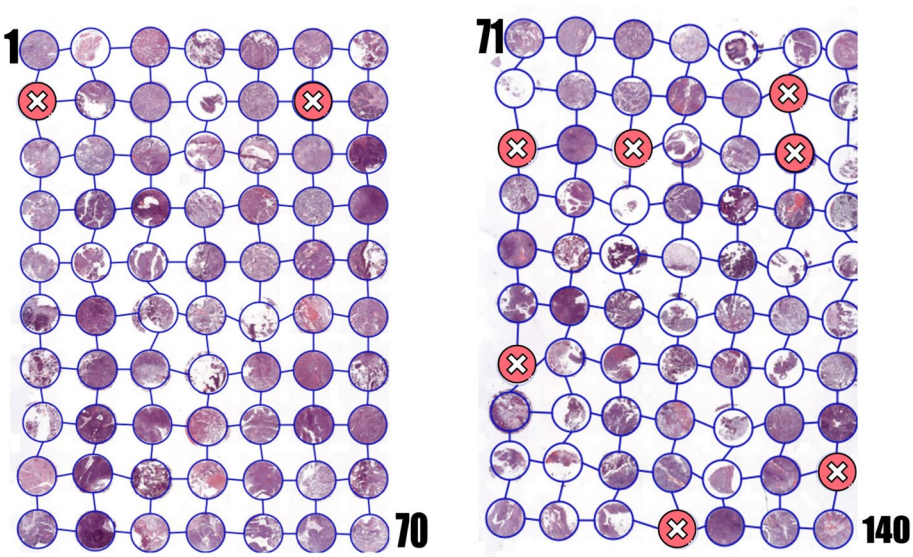

Fig. 1 The two blocks in tissue microarrays comprising 140 cores (9 cores with missing tissue), after QuPath dearraying procedure. There 7 columns and 10 rows. A short script was run to rename the cores from 1 to 140. All the remaining cores had at least 200 tumor cells available for analyses

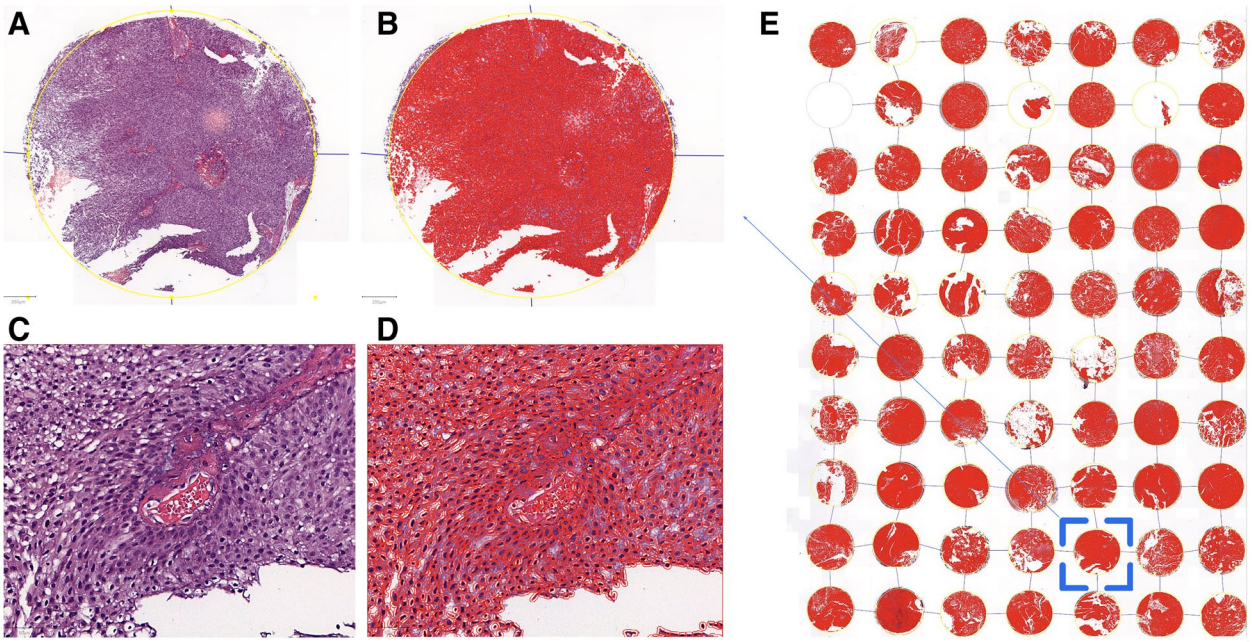

Fig. 2 **A**, core 61 showing high-grade tumor with papillary formation. After Cell Detection, the software shows nuclei detected (red outlines) and cell expansion for each cell (Zoom in in **C** and **D**). For this core, there were 30,204 cellular detections. Panel **E** shows TMA cores 1 to 70 and with the blue square emphasizing #61

A helpful feature for researchers is the ability to distinguish between different cell types depending upon which measurements have been made and present them as a visual heatmap. Figure 3 shows an example (core #34) of a high-grade tumor with the map of the nucleus/cell area ratio output for each cell, and the heatmap for the entire block (cores 1-70) with the same measurement map. For the nucleus/cell area ratio, 1,791,467 cells were measured, with a mean of 0.289 (Std. Dev 0.077). In the figure, yellow represents high N:C ratio, while purple a lower one.

Automated tumor grade recognition

To train the software algorithm in recognizing tumor cells and differentiating cellular classes, 7 cases of high-grade urothelial carcinomas (HG) and 3 cases of low-grade urothelial carcinoma (LG) were annotated, as well as 4 distinct cases of non-tumor tissue (labeled "stroma" and representing non-tumor components such as fibroadipose tissue, desmoplastic stroma, or muscle bundles (Fig. 4, annotations from teaching cores). Within the training set, 4231 tumor cells in 7 high-grade cases, 4906 tumor cells in 4 low-grade cases, and 8023 stroma cells were annotated. A script containing all procedures and steps in detecting and grading was saved, and run on all imported TMA images, thereby automating the detection and across slides, as previously described (Loughrey et al. 2018). This is a particularly helpful feature of the program to increase reproducibility throughout different cases (Kilvaer et al. 2020; Humphries et al. 2018).

Using *Random Trees* classifier based on nuclear and cytoplasmic features (area, perimeter, circularity, caliper, nuclear eccentricity, hematoxylin optical density, eosin optical density (with means and variance) and nuclear to cell area ratio, we tested the algorithm in the first tissue block with 70 cases and made additional training corrections. When the senior pathologist was satisfied, the classification was deployed to all tissue cores, creating a map of each of the thousands of cells. The data was exported to a spreadsheet. For each core, the output showed the number of high-grade, low-grade, and stroma-designated cells (red, blue, and green respectively, see Fig. 5). If high-grade cells were assigned in more than 10% of the tumor cells for each core, the core was designated high-grade by the software. These results were compared to the grade assigned by the pathologist.

There was a good correlation between tumor grade by the pathologist and by QuPath software (Kappa agreement 0.73). Interestingly, the algorithm was almost perfect in diagnosing high-grade tumors (98 out of 99 tumors). The only high-grade tumor classified as low-grade by QuPath was core #14. This core showed 19,552 cells with low-grade prediction and only 2280 with high-grade prediction, thus making it a difficult case to classify. For low-grade carcinomas (by the report and

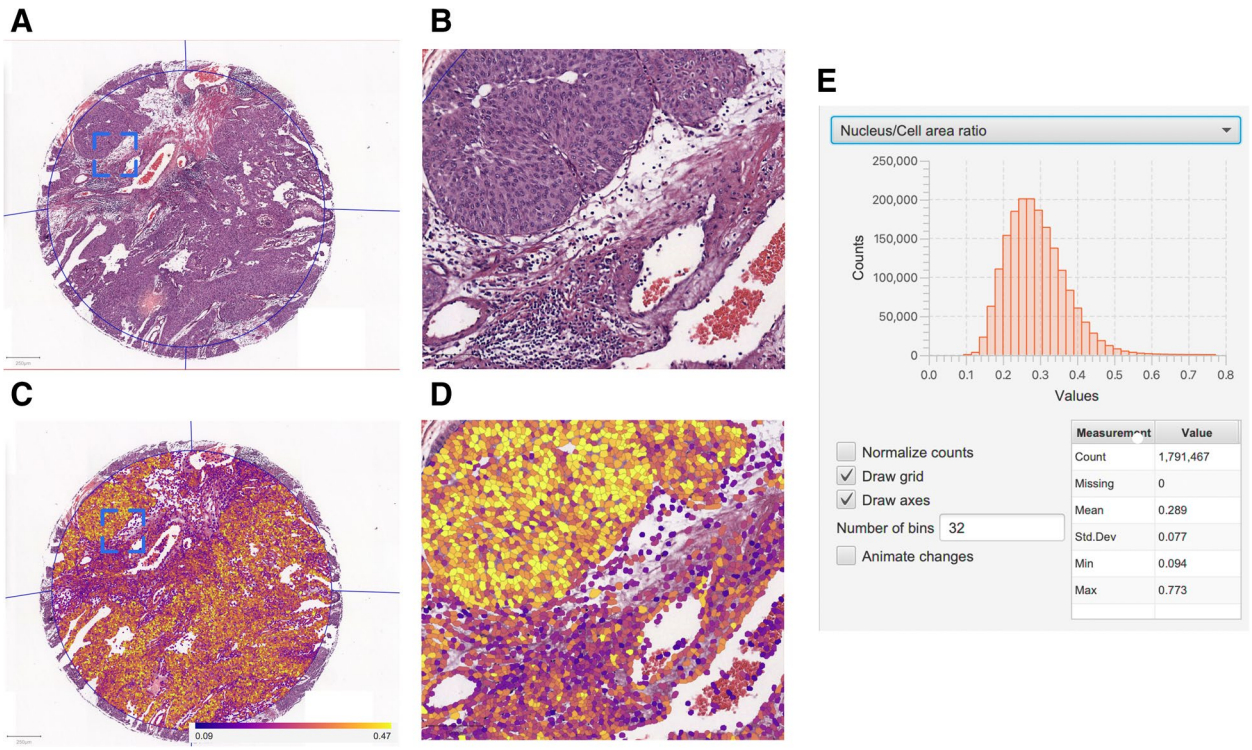

Fig. 3 Nuclear to cell area ratio measurement heatmap. Panel **A** shows the HE section of n invasive papillary high-grade urothelial carcinoma (core #34), with the heatmap colored in each cellular detection (**C**). Towards purple means a lower N:**C** ration and to yellow a higher one. **B** and **D** are zoomed in areas of correspondent **A** and **B** images (blue squares). The heatmap views can be set for any of the output measurements. Panel **E** shows the histogram for all cells in the measurements by normalized values in the software

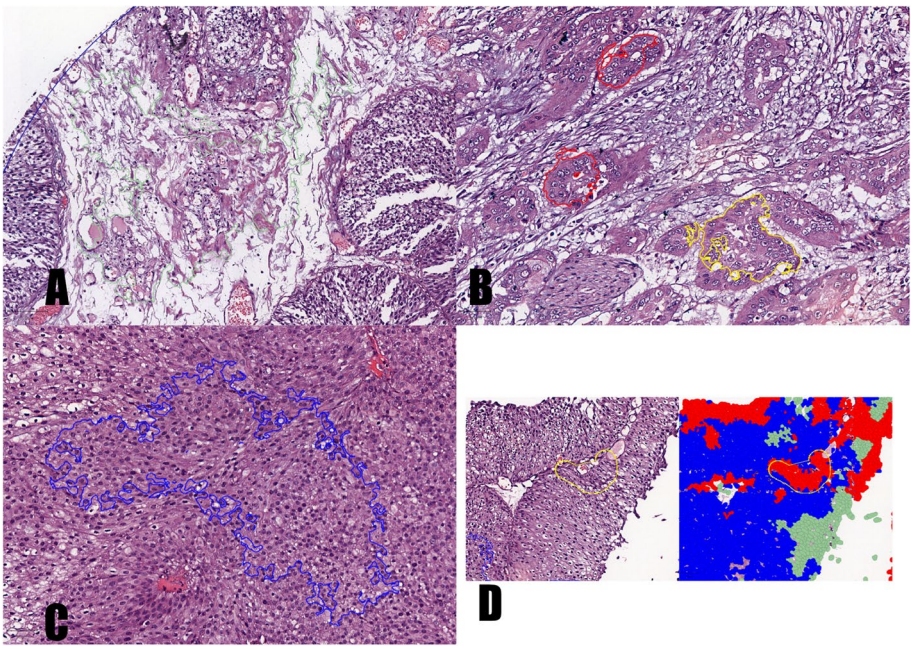

Fig. 4 Training annotations showing stroma (green line, panel **A**), high-grade urothelial carcinoma (green and yellow lines, panel **B**), low-grade urothelial carcinoma, panel **C**. On panel **D**, after the first colored heatmap output, the pathologist corrected the area mistakenly assigned as high grade (red)

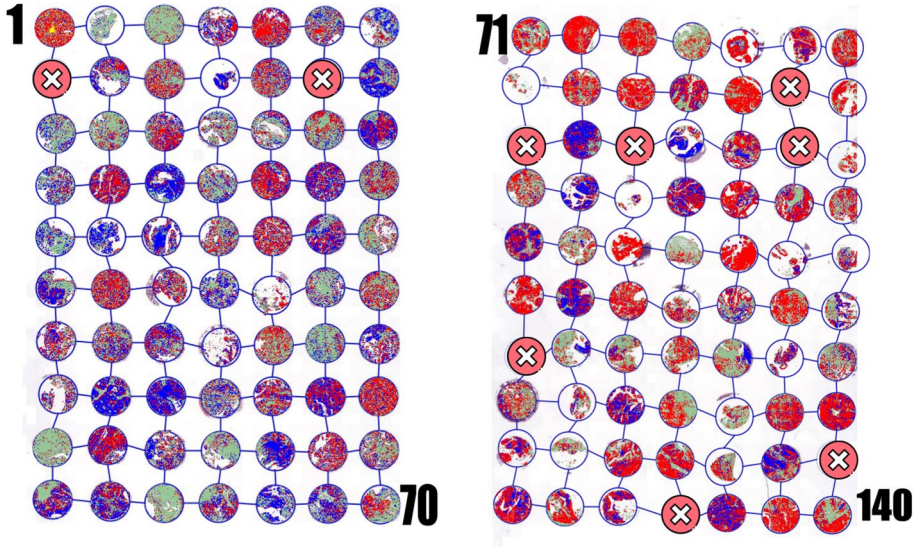

Fig. 5 Final representation of tumor grade heatmap in both TMAS (red is low grade, green is high-grade and stroma (non-tumor tissue) is green)

pathologist), the concordance was not as high. Of the 32 low-grade tumors, 22 were correctly classified as low-grade, but 11 (34%) were diagnosed as high-grade, with the high-grade to the low-grade ratio in these misclassified cases ranging from 0.41 to 0.58. The median ratio for bona fide high-grade carcinomas was 0.59. Some of the reasons the authors list as potential mimickers for high-grade cases are fulguration artifact, nuclear hyperchromasia, folded tissues, and inconsistency in staining.

The software was also able to inform which were the cellular characteristics that were most important to predict between low and high-grade, in this order of importance (in parenthesis, the software provides a number of importance for each variable: (1) nucleus area(0.0805); (2) nuclear hematoxylin optical density mean (0.0705); (3) nucleus circularity (0.0679); (4) nucleus/cell area ratio (0.0645); (5) nucleus perimeter (0.0531); (6) cell area (0.460) and (7) cell minimum caliper (0.0460).

CD-8 and PD-L1 quantification

To validate the ability to count positive cells by immunohistochemistry, the authors visually scored the presence of CD8+ cells within each core and assessed PD-L1 tumor proportion score. These data were compared with the output by QuPath automated *Positive Cell Detection* and *Tissue Detection*. The chosen QuPath output for the CD8 quantification was density (cells/mm2), while we compared both density of PD-L1 positive cells and PD-L1 H-score with the pathologists Tumor Proportion Score (TPS).

The correlation analysis between the software and the pathologist showed that the CD8 marker showed a moderate ($r=0.595$) and statistically significant ($p<0.001$) correlation. The internal consistency of this parameter showed an index of 0.470. The correlation analysis between the software and the pathologist showed that the PD-L1 marker showed a robust ($r=0.834$) and significant ($p < 0.001$) correlation. The internal consistency of this parameter showed a CCI of 0.851. When the H-score was considered for correlation and internal consistency analysis, the correlation was also very strong ($r=0.997$) and significant ($p < 0.001$) and the ICC was 0.919. The agreement showed a statistically significant kappa coefficient ($p < 0.001$) of 0.724. These data are illustrated in Figs. 6 and 7.

Discussion

There is a strong need for accurate quantification of biomarker expression in tissue sections, both in the diagnostic scenario and in the translational research workflows. While molecular techniques have revolutionized tissue-based prediction and prognosis analyses, especially in lung, breast, and colon cancer, the basic surgical pathology sections are cheap, and immunohistochemistry is widely available all over the world.

Digital analyses have overcome some problems of conventional quantitative scoring by pathologists: capacity to evaluate thousands or millions of cells, higher reproducibility in the whole slide scanned images, and speed in evaluating multiple slides or cases in the research setting (Humphries et al. 2018; Ram et al. 2021; Morriss et al.

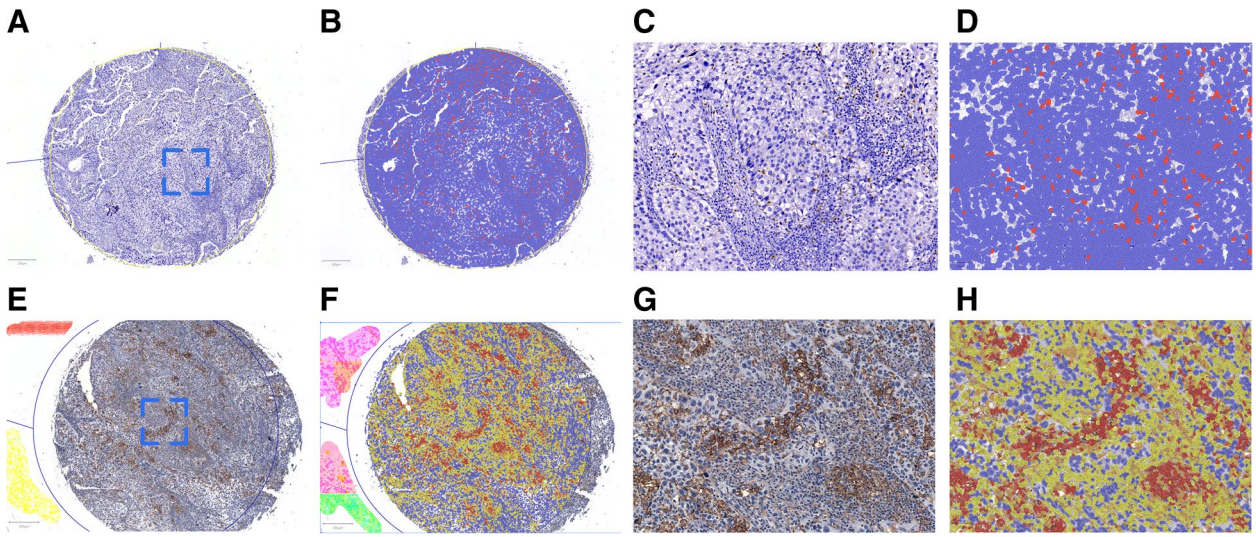

Fig. 6 Original IHC staining for Core #119 showing CD8 in low power (**A**) and colored by the software (red is positive and blue is negative, **B**), and a high-power region (**C** and **D**). Panels **E**, **F**, **G** and **H** are from teh same core stained with PD-L1 (the color outputs are blue for negative, yellow for weak positivity, orange for moderate positivity and red for strong positivity). These scores are used in calculator the H-score

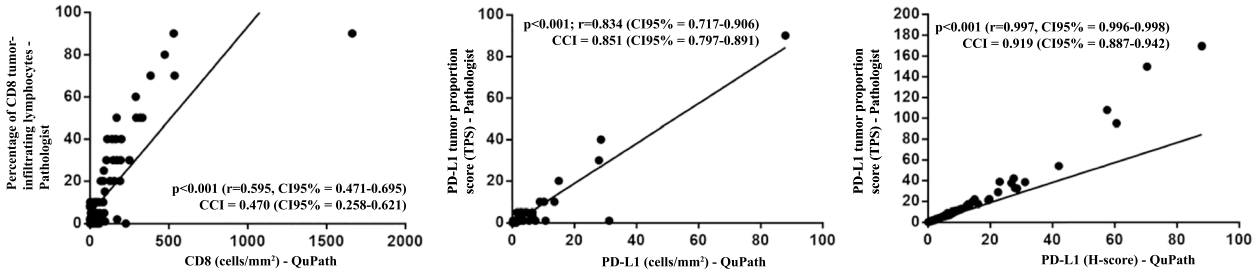

Fig. 7 Correlation analysis and internal consistency of immunoexpression for CD8 and PDL1 using an algorithm. *$p < 0.001$, Spearman correlation

2020; Oh and Mahalingam 2019). Another issue with visual semi-quantitative is the short dynamic range of many systems. The Allred method used in breast cancer assigns scores for intensity (0-3) and percentage of cells (0-5), for example, has a range of 0-8, compared to the output of digital analyses that can yield output of decimals or the more commonly used H-score (0-300) (Meyerholz and Beck 2018; Favresse et al. 2018; Vougiouklakis et al. 2020; Yang et al. 2019). In the past few years, several studies have compared tissue expression evaluated by pathologists and digital algorithms, and more importantly, compared digital with mRNA transcripts and proteomics, underscoring the power of tissue-based protein quantification in research and patient's outcomes (Lu et al. 2022; Junger et al. 2020; Levy-Jurgenson et al. 2020; Officer et al. 2020). Ram et al. quantified protein expression (P-cadherin, PD-L1, and 5T4) and showed a high concordance with both mRNA transcripts and pathologists assigned H-score (Ram et al. 2021). Levy-Jurgenson et al. used a similar approach in an elegant study and used deep-learning algorithms to quantify tumor heterogeneity with excellent correlation with survival by the way of mRNA and miRNA expression from whole-slide images (Levy-Jurgenson et al. 2020).

While other studies have attempted to validate the use of software and compare with real-life pathologist diagnoses, especially using tumor infiltrating lymphocytes and PD-L1 expression (Hendry et al. 2017; Amgad et al. 2020; Corredor et al. 2018; Klauschen et al. 2018; Lu et al. 2019; Udall et al. 2018) the current manuscript is intended to prove something more subtle, but practical, that is, the ability of pathologists with no previous training to use a platform to quantify and compare morphometric data. None of the authors had experience in scripting, coding, or had a background in computer sciences. Bankhead tested the performance of QuPath in colon cancer samples using tissue microarrays and quantified p53, PD-L1, and CD8, elegantly showing the power

of the software in predicting patients' survival (Bankhead et al. 2017).

The utility of assessing PD-L1 in urothelial carcinoma in predicting immune checkpoint inhibitor therapy has been established recently with strong supporting data (van der Heijden et al. 2021; Sotelo et al. 2021; Balar et al. 2017). However, intratumoral heterogeneity in PD-L1 expression may explain some of the controversies and lack of consensus (Weyerer et al. 2021). Five different anti- PD-1 (durvalumab, nivolumab and pembrolizumab) or anti-PD- L1 antibodies (atezolizumab and avelumab) are approved by the FDA for patients with locally advanced urothelial carcinoma. In addition, the complexity of the tumor immune environment can not be established for a positive/negative threshold for PD-L1 expression alone (pembrolizumab: combined positive score, CPS, cut-off = 10%; atezolizumab: immune cell score, cut-off = 5%; durvalumab: tumor cell area/immune cell area algorithm, cut-off = 25%). Therefore, and other predictors such as evaluation, subtyping and quantification of infiltrating lymphocytes have been proposed (Gevaert et al. 2021).

In the current manuscript, we were able to assess CD8+ infiltrating lymphocytes in a moderate-sized cohort of both low-grade and high-grade urothelial carcinoma with good agreement between the pathologist assessment using % of the area (in the cores) and QuPath quantification of the number of cells/mm^2. According to Weyerer et al., little is known about the meaning of differential PD-L1 expression on tumor and immune cells, regarding response to therapy, especially in therapy-naïve patients (Weyerer et al. 2021). Follow-up and treatment information is not available for our cohort, and this may be the great weakness of the paper. However, the main objective was to validate a way of measuring and not establish scientific correlations per se.

Currently, several drug-specific algorithms in immunotherapy are in use worldwide and, while there are attempts at standardization, there is differential sensitivity of the assays, distinct protocols for tumor types, and multiple choices of drugs from the oncologists' point of view. Thus, surgical pathologists face daily challenges in choosing which clone to use and how to score (tumor cells, immune cells, combined scores, etc.) (Udall et al. 2018; Tsao et al. 2017; Doroshow et al. 2021).

Evaluating tumor immune microenvironment is puzzling, since tissue samples are usually limited, and just a static representation of a dynamic and ongoing process. To better understand the interplay between different cells, one approach has been to use multiplex immunohistochemistry with the benefits of utilizing several epitopes to study cellular compositions and relations, concurrently quantifying, and locating tumor microenvironment (Tan et al. 2020; Viratham Pulsawatdi et al. 2020). Xie et al. used a tri-chromogen multiplex IHC and immunofluorescence protocols to characterize immune markers in high-grade urothelial carcinoma and utilized QuPath for immune infiltrate quantification and found good correlation between multiplex and singleplex IHC and between multiplex and manual scoring (Xie et al. 2021).

The methodology in this study included use of TMAs, training the algorithm to recogonize cores with predominantly low and high-grade morphologies, indentify and quantify PD-L1 expression withing tumor cells and calculate CD8 positive cell density within the cores. The use of the random trees classifier protocol to train and test the algorithm have been attempted with results prior (Loughrey et al. 2018; Bankhead et al. 2017). In our cohort of bladder cancer, the morphologic features that allowed best distinction between tumor grade are similar to the ones used routinely by pathologists in assigning malignancy in a subjective way: nuclear size and hyperchromasia (nucleus area and optical density), nuclear circularity, nuclear perimeter, nuclear to cytoplasmic or cell ration and cell area. This feature was complemented with the validation of the calculation of cell density calculated by the identification of DAB positive cells (CD8 and PD-L1), and since PD-L1 is a marker that can mark both tumor as well as immune cells, after tumor cell identification training, the calculation was based only in tumor cell morphology positive for PD-L1. Both had very good correlation with the current state of the art, which is analyses by experienced pathologists either using immune cell density expressed by percentage of area, or percentage of tumor cells positive for PD-L1 such as the Tumor Proportion Score.

In conclusion, we were able to demonstrate the utility of QuPath in identifying and scoring tumor cells and IHC quantification and showed in a step by step manner the directions of using the program in a very basic way. The goal was to illustrate with images an easy way to use this free platform with little experience in programming or even in digital pathology. With the increasing availability and decreasing costs for whole slide scanning hardware, the chance of using a software in both research and diagnostics are hopefully reproducible and will yield large amounts of data and biomarker finding (Deroulers et al. 2013; Huang et al. 2021; Ryu et al. 2019). In the post-pandemic world, we hope and expect that the use of digital images and quantification will be used widely in both clinical diagnoses and research, as these have been validated in many centers (Huang et al. 2021; Hanna et al. 2020).

Abbreviations
CD8: Cluster of differentiation 8; PD-L1: Program death ligand 1; PD-1: Program death 1; IHC: Immunohistochemistry; TPS: Tumor proportion score; CPS: Combined proportion score; HE: Hematoxylin-eosin; TMA: Tissue microarray.

Acknowledgements
Not applicable.

Authors' contributions
AR and FT reviewed all the slides and performed morphological evaluation of histology. FT, AR. JC and PGS participated in the design of the study and performed the statistical analysis. AR, CN, LCM and GV collected the data, reviewed patient information and participated in the laboratory protocols. FT an PGS conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Funding
No external funding was used.

Availability of data and materials
Please contact author for data requests.

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Review Board of UNICHRISTUS, Fortaleza-CE, Brazil.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]ICC, Laboratory of Molecular Biology and Genetics, Fortaleza, Brazil. [2]Argos Laboratory, Av Santos Dumont 5753, #1607, Fortaleza CE, 60175047, Brazil. [3]Department of Pathology, Federal University of Ceara, Fortaleza, Brazil.

Received: 12 February 2022 Accepted: 11 April 2022
Published online: 10 June 2022

References

Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, Sharma A, Kerner JK, Denkert C, Yuan Y, et al. Report on computational assessment of tumor infiltrating lymphocytes from the international Immuno-oncology biomarker working group. NPJ Breast Cancer. 2020;6:16.

Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, Loriot Y, Necchi A, Hoffman-Censits J, Perez-Gracia JL, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017;389:67–76.

Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, et al. QuPath: open source software for digital pathology image analysis. Sci Rep. 2017;7:16878.

Cimino-Mathews A. Novel uses of immunohistochemistry in breast pathology: interpretation and pitfalls. Mod Pathol. 2021;34:62–77.

Corredor G, Wang X, Zhou Y, Lu C, Fu P, Syrigos K, Rimm DL, Yang M, Romero E, Schalper KA, et al. Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer. Clin Cancer Res. 2018;25:1526–34.

Deroulers C, Ameisen D, Badoual M, Gerin C, Granier A, Lartaud M. Analyzing huge pathology images with open source software. Diagn Pathol. 2013;8:92.

Doroshow DB, Bhalla S, Beasley MB, Sholl LM, Kerr KM, Gnjatic S, Wistuba II, Rimm DL, Tsao MS, Hirsch FR. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. Nat Rev Clin Oncol. 2021;18:345–62.

Favresse J, Lardinois B, Chatelain B, Mullier F, Jacqmin H. A reminder of the place of morphology and the H-score in the diagnosis of hemophagocytic lymphohistiocytosis (HLH). Clin Case Rep. 2018;6:527–8.

Gevaert T, Cimadamore A, Montironi R, Eckstein M. PD-L1 testing for Urothelial carcinoma: interchangeability, reliability and future perspectives. Curr Drug Targets. 2021;22:162–70.

Gray RT, Cantwell MM, Coleman HG, Loughrey MB, Bankhead P, McQuaid S, O'Neill RF, Arthur K, Bingham V, McGready C, et al. Evaluation of PTGS2 expression, PIK3CA mutation, aspirin use and Colon Cancer survival in a population-based cohort study. Clin Transl Gastroenterol. 2017;8:e91.

Guo H, Ding Q, Gong Y, Gilcrease MZ, Zhao M, Zhao J, Sui D, Wu Y, Chen H, Liu H, et al. Comparison of three scoring methods using the FDA-approved 22C3 immunohistochemistry assay to evaluate PD-L1 expression in breast cancer and their association with clinicopathologic factors. Breast Cancer Res. 2020;22:69.

Gurgel DC, Wong DVT, Bandeira AM, Pereira JFB, Gomes-Filho JV, Pereira AC, Barros Silva PG, Tavora FRF, Pereira AF, Lima-Junior RCP, Almeida PRC. Cytoplasmic CCR7 (CCR7c) immunoexpression is associated with local tumor recurrence in triple-negative breast cancer. Pathol Res Pract. 2020;216:153265.

Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, Busam KJ, Sauter JL, Brogi E, Tan LK, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol. 2020;33:2115–27.

Hardy LB, Fitzgibbons PL, Goldsmith JD, Eisen RN, Beasley MB, Souers RJ, Nakhleh RE. Immunohistochemistry validation procedures and practices: a College of American Pathologists survey of 727 laboratories. Arch Pathol Lab Med. 2013;137:19–25.

Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol. 1999;17:1474–81.

Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, Christie M, van de Vijver K, Estrada MV, Gonzalez-Ericsson PI, et al. Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the international Immunooncology biomarkers working group: part 1: assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. Adv Anat Pathol. 2017;24:235–51.

Huang W, Randhawa R, Jain P, Iczkowski KA, Hu R, Hubbard S, Eickhoff J, Basu H, Roy R. Development and validation of an artificial intelligence-powered platform for prostate Cancer grading and quantification. JAMA Netw Open. 2021;4:e2132554.

Humphries MP, Hynes S, Bingham V, Cougot D, James J, Patel-Socha F, Parkes EE, Blayney JK, O'Rorke MA, Irwin GW, et al. Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER−/HER2+ breast Cancer. J Oncol. 2018;2018:2937012.

Humphries MP, Maxwell P, Salto-Tellez M. QuPath: The global impact of an open source digital pathology system. Comput Struct Biotechnol J. 2021;19:852–9.

Ilyas M, Grabsch H, Ellis IO, Womack C, Brown R, Berney D, Fennell D, Salto-Tellez M, Jenkins M, Landberg G, et al. Guidelines and considerations for conducting experiments using tissue microarrays. Histopathology. 2013;62:827–39.

Junger H, Dobi D, Chen A, Lee L, Vasquez JJ, Tang Q, Laszik ZG. Novel in situ hybridization and multiplex immunofluorescence technology combined with whole-slide digital image analysis in kidney transplantation. J Histochem Cytochem. 2020;68:445–59.

Kilvaer TK, Paulsen EE, Andersen S, Rakaee M, Bremnes RM, Busund LR, Donnem T. Digitally quantified CD8+ cells: the best candidate marker for an immune cell score in non-small cell lung cancer? Carcinogenesis. 2020;41:1671–81.

Klauschen F, Müller KR, Binder A, Bockmayr M, Hägele M, Seegerer P, Wienert S, Pruneri G, de Maria S, Badve S, et al. Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning. Semin Cancer Biol. 2018;52:151–7.

Levy-Jurgenson A, Tekpli X, Kristensen VN, Yakhini Z. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. Sci Rep. 2020;10:18802.
Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AMB, Gray RT, McQuaid S, Dunne PD, Hamilton PW, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. Histopathology. 2018;73:327–38.
Lu S, Stein JE, Rimm DL, Wang DW, Bell JM, Johnson DB, Sosman JA, Schalper KA, Anders RA, Wang H, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade. JAMA Oncol. 2019;5:1195.
Lu X, Yu P, Tao H, Chen Y, Lu J. Correlation between TOPA2A gene expression and the number of CD4(+) T cells in hepatocellular carcinoma and its clinical prognostic significance. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi. 2022;38:24–31.
McCarty KS Jr, Szabo E, Flowers JL, Cox EB, Leight GS, Miller L, Konrath J, Soper JT, Budwit DA, Creasman WT, et al. Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors. Cancer Res. 1986;46:4244s–8s.
McGinnis LM, Ibarra-Lopez V, Rost S, Ziai J. Clinical and research applications of multiplexed immunohistochemistry and in situ hybridization. J Pathol. 2021;254:405–17.
Meyerholz DK, Beck AP. Principles and approaches for reproducible scoring of tissue stains in research. Lab Investig. 2018;98:844–55.
Morais CE, Gurgel DC, Teixeira AC, Mattos TVA, Silva A, Tavora F. Prevalence of ERG expression and PTEN loss in a Brazilian prostate cancer cohort. Braz J Med Biol Res. 2019;52:e8483.
Morriss NJ, Conley GM, Ospina SM, Meehan WP, Qiu J, Mannix R. Automated Quantification of Immunohistochemical Staining of Large Animal Brain Tissue Using QuPath Software. Neuroscience. 2020;429:235–44 Elsevier Ltd.
Officer LK, Andreou KE, Teodósio AV, He Z, Le Quesne JP. Automated co-in situ hybridization and immunofluorescence using archival tumor tissue. Methods Mol Biol. 2020;2148:245–56.
Oh KS, Mahalingam M. Immunohistochemistry as a genetic surrogate in Dermatopathology: pearls and pitfalls. Adv Anat Pathol. 2019;26:390–420.
Ozbudak IH, Shilo K, Tavora F, Rassaei N, Chu WS, Fukuoka J, Jen J, Travis WD, Franks TJ. Glucose transporter-1 in pulmonary neuroendocrine carcinomas: expression and survival analysis. Mod Pathol. 2009;22:633–8.
Ram S, Vizcarra P, Whalen P, Deng S, Painter CL, Jackson-Fisher A, Pirie-Shepherd S, Xia X, Powell EL. Pixelwise H-score: a novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. PLoS One. 2021;16:e0245638.
Ross J, Li G, Yang XJ. Application and pitfalls of immunohistochemistry in diagnosis of challenging genitourinary cases. Arch Pathol Lab Med. 2020;144:290–304.
Ryu HS, Jin MS, Park JH, Lee S, Cho J, Oh S, Kwak TY, Woo JI, Mun Y, Kim SW, et al. Automated Gleason scoring and tumor quantification in prostate Core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. Cancers (Basel). 2019;11:1860.
Satturwar S, Malenie R, Sutton A, Dai D, Aly FZ. Validation of immunohistochemical tests performed on cytology cell block material: practical application of the College of American Pathologists' guidelines. Cytojournal. 2019;16:6.
Shah AA, Bourne TD, Murali R. BAP1 protein loss by immunohistochemistry: a potentially useful tool for prognostic prediction in patients with uveal melanoma. Pathology. 2013;45:651–6.
Shelton J, Purgina BM, Cipriani NA, Dupont WD, Plummer D, Lewis JS Jr. p16 immunohistochemistry in oropharyngeal squamous cell carcinoma: a comparison of antibody clones using patient outcomes and high-risk human papillomavirus RNA status. Mod Pathol. 2017;30:1194–203.
Sotelo M, Alonso-Gordoa T, Gajate P, Gallardo E, Morales-Barrera R, Perez-Gracia JL, Puente J, Sanchez P, Castellano D, Duran I. Atezolizumab in locally advanced or metastatic urothelial cancer: a pooled analysis from the Spanish patients of the IMvigor 210 cohort 2 and 211 studies. Clin Transl Oncol. 2021;23:882–91.
Tan WCC, Nerurkar SN, Cai HY, Ng HHM, Wu D, Wee YTF, Lim JCT, Yeong J, Lim TKH. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Commun (Lond). 2020;40:135–53.
Tsao M, Kerr K, Yatabe Y, Hirsch FR. PL 03.03 blueprint 2: PD-L1 immunohistochemistry comparability study in real-life, clinical samples. J Thorac Oncol. 2017;12:S1606.
Udall M, Rizzo M, Kenny J, Doherty J, Dahm S, Robbins P, Faulkner E. PD-L1 diagnostic tests: a systematic literature review of scoring algorithms and test-validation metrics. Diagn Pathol. 2018;13:12.
van der Heijden MS, Loriot Y, Duran I, Ravaud A, Retz M, Vogelzang NJ, Nelson B, Wang J, Shen X, Powles T. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic Urothelial carcinoma: a long-term overall survival and safety update from the phase 3 IMvigor211 clinical trial. Eur Urol. 2021;80:7–11.
Viratham Pulsawatdi A, Craig SG, Bingham V, McCombe K, Humphries MP, Senevirathne S, Richman SD, Quirke P, Campo L, Domingo E, et al. A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. Mol Oncol. 2020;14:2384–402.
Vougiouklakis T, Belovarac BJ, Lytle A, Chiriboga L, Ozerdem U. The diagnostic utility of EZH2 H-score and Ki-67 index in non-invasive breast apocrine lesions. Pathol Res Pract. 2020;216:153041.
Weyerer V, Strissel PL, Strick R, Sikic D, Geppert CI, Bertz S, Lange F, Taubert H, Wach S, Breyer J, et al. Integration of spatial PD-L1 expression with the tumor immune microenvironment outperforms standard PD-L1 scoring in outcome prediction of Urothelial Cancer patients. Cancers (Basel). 2021;13:2327.
Xie Y, Olkhov-Mitsel E, Alminawi S, Slodkowska E, Downes MR. Development of a multiplex immuno-oncology biomarker and digital pathology workflow for assessment of urothelial carcinoma. Pathol Res Pract. 2021;226:153607.
Yang Y, Xiao M, Song Y, Tang Y, Luo T, Yang S, He W, Cheng Q, Ma L, Zhang Y, et al. H-score of 11beta-hydroxylase and aldosterone synthase in the histopathological diagnosis of adrenocortical tumors. Endocrine. 2019;65:683–91.

Publisher's Note