ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

IOTECH IMMUNO-ONCOLOGY TECHNOLOGY

REVIEW

Development and applications of computer image analysis algorithms for scoring of PD-L1 immunohistochemistry

L. J. Inge* **& E. Dennis**

Roche Tissue Diagnostics, Tucson, USA

Available online 11 May 2020

Immune checkpoint inhibitors targeting programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) have rapidly become integral to standard-of-care therapy for non-small cell lung cancer and other cancers. Immunohistochemical (IHC) staining of PD-L1 is currently the accepted and approved diagnostic assay for selecting patients for PD-L1/PD-1 axis therapies in certain indications. However, the inherent biological complexity of PD-L1 and the availability of several PD-L1 assays — each with different detection systems, platforms, scoring algorithms and cut-offs — have created challenges to ensure reliable and reproducible results based on subjective visual assessment by pathologists. The increasing adoption of computer technologies into the daily workflow of pathology provides an opportunity to leverage these tools towards improving the clinical value of PD-L1 IHC assays. This review describes several image analysis software programs of computer-aided PD-L1 scoring in the hope of driving further discussion and technological advancement in digital pathology and artificial intelligence approaches, particularly as precision medicine evolves to encompass accurate simultaneous assessment of multiple features of cancer cells and their interactions with the tumor microenvironment.

Key words: PD-L1, non-small cell lung cancer, immunohistochemistry, machine learning, image analysis, digital pathology

BACKGROUND

Basic and translational studies have defined programmed cell death 1 (PD-1)/programmed cell death ligand 1 (PD-L1) signaling as a key regulator of T-cell and immune responses, which led to their recognition as therapeutic targets for cancer.[1] Clinical trial data have demonstrated that inhibitory antibodies against PD-1 or PD-L1 may produce durable, efficacious responses as either monotherapy or in combination regimens.[2–6] Side effects from these treatments are manageable, although serious adverse events, particularly excessive immune activation, can occur.[7,8] PD-1/PD-L1-targeted therapies have thus become integrated into standard-of-care treatment for several tumor types (Table 1), with continued expansion underway.

The increasing use of PD-1/PD-L1 inhibitors has highlighted the key role of biomarkers in identifying patient populations who will gain greatest benefit from these treatments. Currently, immunohistochemical (IHC) analysis of PD-L1 expression is the most widely used test for assessing patient eligibility for anti-PD-1/PD-L1 treatments in certain indications[9,10] (Table 1), and is approved as companion or complementary diagnostics by health authorities globally. Studies show that patients with a higher percentage of PD-L1-positive cells [e.g. tumor cells (TCs); immune cells (ICs), including macrophages, myeloid and lymphocytes] may be more likely to respond to treatment compared with patients whose tumors display lower amounts of PD-L1 positivity.[2–6,9–12] A variety of different requirements and recommendations exist for PD-L1 IHC testing in different tumor types (Table 1). Regardless of the assay, all PD-L1 IHC tests require a pathologist to evaluate the staining patterns and determine the percentage of PD-L1-positive TCs (%TC) and/or PD-L1-positive ICs (%IC) within the viable tumor area to make a determination on patient eligibility for a particular therapy in a particular indication.

Several factors can create challenges for pathologists scoring PD-L1 in tumors. These include: (i) the temporal and spatial heterogeneity of PD-L1 as an adaptive, physiological biomarker; (ii) difficulty scoring poorly circumscribed and/or heterogeneous tumors; (iii) pre-analytical variables that affect staining quality (e.g. fixation time, antigen stability) and obscuring endogenous material (e.g. anthracotic pigment); (iv) the potential exclusion of PD-L1-positive ICs (specific to %TC scoring algorithms); (v) misinterpretation of

*Correspondence to: Landon J. Inge, Medical Affairs Manager, Global Medical Affairs, Roche Tissue Diagnostics, 1910 Innovation Park Dr., Tucson, AZ 85755, USA. Tel: 520-336-6454

E-mail: landon.inge@roche.com (L.J. Inge).

Table 1. Disease indications, testing requirements and approved diagnostic scoring algorithms for approved programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) therapies.

ICI	Current indications	Testing requirements	Scoring cut-off	Indications
Pembrolizumab	NSCLC	PD-L1 IHC is required (companion)	≥1% TPS	1L for mNSCLC and unresectable stage III NSCLC
	HNSCC		≥1% CPS	2L for mNSCLC
	UC		≥10% CPS	1L for unresectable HNSCC
	G/GEJ Adenocarcinoma			2L for recurrent locally advanced or metastatic G/GEJ adenocarcinoma
	ESCC			2L for recurrent or metastatic cervical cancer
	Cervical cancer			1L for locally advanced and metastatic UC
	MSI-H[a]			2L for recurrent locally advanced or metastatic ESCC
	Melanoma	Testing is not required		
	SCLC			
	cHL			
	PMBCL			
	HCC			
	MCC			
	RCC			
Nivolumab	NSCLC	PD-L1 IHC is suggested (complementary)	≥1, 5 or 10%	2L for non-squamous NSCLC
	HNSCC		≥1%	2L for HNSCC
	UC			2L for UC
	MSI-H[a]			
	Cervical cancer	Testing is not required		
	HCC			
	cHL			
	RCC			
	SCLC			
	Melanoma			
Durvalumab	NSCLC	PD-L1 IHC is suggested (complementary)	≥1% TCs	2L for unresectable NSCLC that has not progressed on chemoradiation
	UC		≥25% TCs or ≥25% ICs and ICP>1% or ICP = 1% and IC = 100%	2L for locally advanced or metastatic UC
Atezolizumab	UC	PD-L1 IHC is required (companion)	≥1% ICs	1L for TNBC
	TNBC		≥5% ICs	1L for UC
	NSCLC	PD-L1 IHC is suggested (complementary)	≥50% TCs or ≥10% ICs	1L for mNSCLC
			≥1% TCs or ≥1% ICs	2L for mNSCLC
	SCLC	Testing is not required		

cHL, classical Hodgkin lymphoma; CPS, combined positive score; ESCC, squamous cell carcinoma of the esophagus; G/GEJ, gastric or gastroesophageal junction; HCC, hepatocellular carcinoma; HNSCC, head and neck squamous cell cancer; IC, immune cell; ICP, immune cells present; IHC, immunohistochemistry; MCC, Merkel cell carcinoma; MSI-H, microsatellite instability-high cancer; mNSCLC, metastatic non-small cell lung cancer; NSCLC, non-small cell lung cancer; PMBCL, primary mediastinal large B-cell lymphoma; RCC, renal cell carcinoma; SCLC, small cell lung cancer; TC, tumour cell; TNBC, triple-negative breast cancer; TPS, tumor proportion score; UC, urothelial carcinoma; 1L, first-line; 2L, second- line.

[a] Only a positive test for MSI-H is required.

staining (e.g. cytoplasmic); and (vi) non-specific staining (e.g. necrotic tissue).[9,13] Moreover, scoring of borderline cases can also be challenging.[14,15] Although studies indicate that agreement is achievable between pathologists for PD-L1 IHC scoring,[15–18] a recent multi-institutional study of PD-L1 assessment found that agreement decreases as the number of observers increases.[14,15,19] Educational programs that provide pathologists with training to develop familiarity and expertise in PD-L1 scoring have been shown to mitigate scoring issues.[17,18,20] However, faced with growing work demands and ongoing evolution of the PD-L1 diagnostic landscape, development of new tools that can aid pathologists in their daily workflow[14,15] would have medical value.

As in other clinical specialties (e.g. radiology, ophthalmology[21,22]), computers are impacting the clinical workflow of pathology. Technologies now enable the capture and storage of high-resolution whole slide images (WSIs) from glass slides that can be reviewed by a pathologist using a computer and/or a mobile device, negating the need for a microscope.[23] Computer-aided analysis of WSIs is used to aid and/or automate repetitive and basic tasks, such as counting cells and two-dimensional measurements, that are part of a pathologist's daily work routine.[24] WSI systems (e.g. scanner, image management system, computer workstation) have been approved for primary diagnosis by the US Food and Drug Administration and European Union regulatory agencies. Computer image analysis (IA) algorithms are currently approved for use in the clinical diagnosis of HER2, Ki67 and estrogen receptor/progesterone receptor in breast cancer.[25,26]

With the wider adoption of WSIs, there is an increase in the application of artificial intelligence (AI) and machine learning (ML) to utilize the highly complex visual information residing in WSIs to perform diagnostic tasks. AI is a broad area of computer science focused on developing computer programs that can perform tasks and solve problems that require human intelligence. ML, defined by Arthur Samuel in 1959 as a "field of study that gives computers the ability to learn without being explicitly programmed",[27] exists as a branch of AI. Although a full discussion of AI/ML and applications in pathology are outwith the scope of this review (see elsewhere for more in-depth discussion[28–31]), it is valuable to appreciate the

different approaches of ML, specifically supervised, unsupervised and semi-supervised learning. Collectively, these approaches utilize data inputted into the algorithm for the learning task, but differ in how the input and output data are defined. In supervised learning, the algorithm is given highly specific, annotated data and specific classifiers with which to define both the input and output data. In unsupervised learning, the algorithm is given raw input data and no classifiers for defining the output data. Semi-supervised learning attempts to bridge the gap between the two by providing annotated data to the algorithm for training but no classifiers to define the output data. Although the two terms are sometimes used interchangeably, deep learning is an evolution of ML, taking inspiration from the human brain in the design of the learning (i.e. supervised, unsupervised, semi-supervised) tasks. Each method has strengths and weaknesses, and application is dependent upon the type and amount of data. In pathology, supervised learning has been the more commonly used method, utilizing images manually annotated by pathologists. However, this can be problematic as the requirement of large sets of images expertly annotated by pathologists can be difficult to source. As such, ML approaches that allow for less extensively annotated data, such as images segregated into basic classes (e.g. tumor, normal or tumor, stroma), have been applied. ML also allows for computers to capture and incorporate subvisual morphometric features into algorithms that may be more accurate. These efforts have produced IA algorithms that can perform complex tasks, including classification of breast cancer stroma,[32,33] identification of lymph node metastases,[34] prediction of patient prognosis in non-small cell lung cancer (NSCLC),[35] and detection of microsatellite instability in gastrointestinal cancer.[36] The ability of these IA algorithms to perform these tasks accurately and successfully supports both the continued development and application of IA algorithms to aid pathologists in analytical tasks.

APPLICATION OF IA ALGORITHMS IN THE ANALYSIS OF PD-L1

IA algorithms are already being used to investigate the role of PD-L1, providing considerable insight into its expression and heterogeneity within the tumor microenvironment.[37] Moreover, applications of IA algorithms, in combination with chromogenic and immunofluorescent IHC, are enabling in-depth quantitative analysis of the broader tumor microenvironment.[38–41] Due to the value of IA algorithms as a research tool, leveraging IA algorithms as a clinical tool for pathologists is an obvious progression. Indeed, several studies have described PD-L1 IA algorithms that show the potential for IA as a clinical tool in PD-L1 evaluation (Table 2). Koelzer et al. used the HALO™ IA software (Indica Labs, Albuquerque, NM) to develop a PD-L1 scoring algorithm for melanomas.[42] The authors trained the algorithm using a supervised learning approach termed 'Random Forest'. Random Forest is a data classification system that utilizes many different decision trees that work together to classify objects within an image. As predictions made by each individual decision tree share little correlation, the collective decision trees (i.e. the 'random' in Random Forest) function as a committee and provide a much more accurate prediction than each individual prediction. Furthermore, this collective approach may reduce errors, as while one decision tree may be wrong, other trees will be right. Using this approach, Koelzer et al. trained the algorithm to recognize membrane PD-L1 staining in TCs, recognize unstained TCs, exclude ICs, and produce a %TC score after manual annotation of the tumor area by the pathologist. This IA algorithm was highly concordant with the %TC scores generated by two pathologists using

Table 2. Overview of selected programmed cell death ligand 1 (PD-L1) image analysis (IA) algorithms.

Author	ML method	Tumor type	Scoring type	Sample dataset	Relevant data	Reference
Koelzer et al.	Random forest/ supervised learning	Melanoma	%TC	69 samples of melanoma	Pearson correlation coefficient ($r = 0.97$, $P < 0.0001$) between pathologist and IA	42
Kim et al.	Supervised learning	Gastric cancer	CPS	39 patients with clinical response to pembrolizumab	Correlation of PD-L1 positivity with patient (RFS) outcome [HR 0.536 (95% CI 0.316 −0.94), $P = 0.0294$]	43
Humphries et al.	Supervised learning	TNBC	% positive PD-L1	90 samples with clinical outcome	Correlation of PD-L1 positivity with patient (RFS) outcome [HR 0.536 (95% CI 0.316 −0.94), $P = 0.0294$]	44
Kapil et al	GAN/semi-supervised learning	NSCLC (biopsies)	TPS[a]	270 needle core biopsies; 60 slides used for concordance of manual to IA scores	IA scoring concordance with visual scores (OPA = 0.88, NPA = 0.88, PPA = 0.85; Lin's CCC = 0.94; Pearson CCC = 0.95)	45
Taylor et al.	Supervised learning with feedback loop	NSCLC	%TC, %IC	230 cases	Concordance (Lin's CCC) of IA with three pathologists (%TC = 0.81, 0.78, 0.68; %IC = 0.62, 0.53, 0.88)	46

%IC, percentage of PD-L1-positive immune cells; %TC, percentage of PD-L1-positive tumour cells; CCC, concordance correlation coefficient; CI, confidence interval; CPS, combined positive score; GAN, generative adversarial network; HR, hazard ratio; ML, machine learning; NPA, negative percent agreement; NSCLC, non-small cell lung cancer; OPA, overall percent agreement; PPA, positive percent agreement; RFS, relapse-free survival; TNBC, triple-negative breast cancer; TPS, tumor proportion score.
[a] TPS calculated from positive and negative pixels.

conventional light microscopy (Pearson correlation coefficient $r = 0.97$, $P < 0.0001$), and improved interobserver agreement between pathologists to near-perfect agreement. Kim et al. used the Aperio IHC membrane IA algorithm (Leica Biosystems, Wetzlar) to quantify PD-L1 staining across all cell types in 39 cases of gastric cancer.[43] Notably, the Aperio IHC membrane algorithm is not designed to score PD-L1 specifically, and instead quantifies DAB staining intensity. As such, the authors were required to perform PD-L1 analysis through manual annotation and computation to generate the PD-L1 combined positive TC score. Even without computer-aided selection and computation of the PD-L1 score, the authors showed that application of the IHC membrane algorithm as a supportive tool in PD-L1 evaluation provided PD-L1 IA scores which were concordant (Fisher's exact test 84.6%, $P < 0.05$) with manual scoring. Furthermore, PD-L1 IA scores were comparable with manual scoring in predicting patient response to pembrolizumab (Delong's test for two correlated receiver operating characteristic curves, $P = 0.186$). More recently, Humphries et al. used QuPath (https://qupath.github.io/; developed by the Queen's University Belfast, Belfast, Northern Ireland, UK), an open source digital pathology IA software platform, to perform PD-L1 analysis in triple-negative breast cancer.[44] Relying on the generic tool for quantifying DAB intensity, the authors further refined the algorithm using reference cell lines and manual selection of tumor and stromal tissue to determine PD-L1 positivity. To evaluate the IA algorithm, the authors assessed the correlation of the IA score of PD-L1 positivity with clinical outcome, instead of comparison with manual scoring by a pathologist. Using the clinical 1% cut-off for PD-L1 positivity in triple-negative breast cancer, the authors demonstrated correlation between IA scores >1% and increased relapse-free survival.

Counter to requiring manual annotation to achieve successful PD-L1 analysis, Kapil et al. developed a fully automated IA algorithm for PD-L1 IHC staining of NSCLC needle biopsies.[45] To develop the algorithm, the authors chose to apply a novel approach to machine learning, termed 'generative adversarial networks' (GANs). GANs use two competing neural networks (the generator and discriminator) to classify and interpret data. The generator produces data that the discriminator evaluates against the ground truth data. Due to the opposing goals of each network, GANs are able to train each other and generate highly accurate data using small, minimally annotated datasets. Using the GAN approach, the authors were able to develop an automated PD-L1 scoring algorithm with a minimal ground truth dataset. PD-L1 scores generated from the GAN-trained algorithm showed higher concordance with manual scoring by a pathologist compared with scores from algorithms trained using supervised or semi-supervised learning methods.[45]

The Optra PD-L1 IA algorithm (OptraSCAN Inc, San Jose, CA) was developed using supervised learning and a feedback loop process, wherein the algorithm is further trained by assessing its performance on unknown datasets, combined with data evaluation by a board-certified pathologist with a feedback loop to the algorithm. The benefit of the feedback loop is in its ability to allow the algorithm to further improve the quality of the data produced. The PD-L1 algorithm produced through this process was validated across three institutions for scoring both PD-L1-positive TCs and ICs in NSCLC.[46] PD-L1 scoring between the IA algorithm and three pathologists was concordant for TCs [Lin's concordance correlation coefficient (CCC) = 0.81, 0.78 and 0.68]; however, concordance for ICs was less between the algorithm and the pathologists (Lin's CCC = 0.62, 0.53 and 0.88). Notably, concordance of manual scoring between pathologists was strong for TCs but not for ICs (Lin's CCC = 0.56, 0.48 and 0.48).

Discerning between positive TCs and ICs can be challenging in PD-L1 IHC interpretation; for example, distinguishing TCs from macrophage staining, or determining IC staining where there is strong TC staining. Both Koelzer et al. and Taylor et al. developed IA algorithms that identify both TCs and ICs and either exclude them from scoring[42] or provide a PD-L1-positive score for ICs.[46] The ability of these algorithms to identify ICs correctly is increasingly relevant. As noted in Table 1, the need to exclude or include PD-L1-positive ICs is dependent upon the assay used, and the PD-1/PD-L1 inhibitor and tumor type being assessed. In some tumor types (e.g. triple-negative breast cancer and urothelial carcinoma), scoring of PD-L1-positive ICs alone forms the diagnostic algorithm for the VENTANA PD-L1 (SP142) assay (F. Hoffmann-La Roche AG, Basel). However, unlike visual PD-L1 scoring of TCs, visual scoring of PD-L1-positive ICs can be more challenging, especially if pathologists are not trained on that specific assay.[16,47,48] Based on the ability of these IA algorithms to classify heterogeneous and spatially mixed cell populations of TCs and ICs, it is possible that IA algorithms could function as an orthogonal method for PD-L1 testing, thus alleviating concerns about accurate scoring of ICs,[16,47,48] particularly for borderline and challenging cases.

Despite the widespread adoption of PD-L1 IHC as a clinical decision tool, questions remain regarding its potential limitations. Based on multiple observations that some patients with low PD-L1 levels demonstrate a clinical response to PD-1/PD-L1 therapy while some patients with high PD-L1 levels fail to show any clinical benefit, a variety of investigations towards alternative diagnostic tests have been undertaken. Multiplex IHC (mIHC) visualizes multiple protein targets within a single section,[49,50] and is a logical application for IA algorithms to provide diagnostic information. Indeed, the application of IA algorithms with manual mIHC has been used successfully to characterize the tumor microenvironment in correlation with patient outcome.[51–53] Automated mIHC with IA algorithms have also been used successfully.[41,54–56] For example, Developer XD™ 2.7 (Definiens AG, München) was applied to clinical trial samples from patients with NSCLC treated with durvalumab to develop a CD8+/PD-L1+ signature that was more predictive of patient outcome than manual assessment of %TC PD-L1.[57] Correlation was found between the

CD8+/PD-L1+ signature and improved median overall survival time (24.2 months), compared with PD-L1-alone IHC (15.5 months). Although a definitive correlation with patient outcome was not shown, Silva et al. used cTA™ IA (Flagship Biosciences, Inc., Westminster, CO) to characterize the NSCLC tumor microenvironment and show that NSCLC tumors displaying PD-L1 positivity >50% had increased numbers of macrophages (CD68+) and T-cells (CD8, FoxP3, Granzyme B) residing in the tumor nests.[58] Although these studies described analysis of chromogenic serial sections stained with single antibodies, mIHC in a single section with IA has potential to provide even more diagnostic value.[49,54] In fact, a recent meta-analysis of diagnostic tests for PD-1/PD-L1 therapies showed that mIHC had better predictive value [area under the curve (AUC) = 0.79] than PD-L1 IHC (AUC = 0.65, $P < 0.001$) and nucleic-acid-based assays [gene expression profiling (AUC = 0.65, $P = 0.003$), tumor mutational burden (AUC = 0.69, $P = 0.049$)].[50]

FUTURE DIRECTIONS AND CONCLUSIONS

The studies highlighted in this review demonstrate the potential of WSI and IA algorithms to aid pathologists in improving the accuracy and reproducibility of PD-L1 scoring. The studies described in this review highlight approaches to developing algorithms that can produce scores which are consistent with manual scoring, and provide additional diagnostic information. Although this review has highlighted the potential benefits of IA algorithms in improving the reproducibility of PD-L1 scoring, caution is warranted. Critically, none of the IA algorithms described in this review are approved for clinical diagnosis, and they have been tested using small sample sets and/or in small studies. Furthermore, although these algorithms have, within these limited datasets, shown concordance with manual evaluation, it remains to be seen whether these AI approaches are capable of demonstrating a high level of performance in real-world settings. Before discussions about clinical adoption of these PD-L1 algorithms can begin, it is critical to test these algorithms at multiple sites using larger sample cohorts whose clinical outcome is known. Indeed, as demonstrated by Humphries et al.[44] and Kim et al.,[43] the incorporation of clinical outcome as one of the benchmarks for IA algorithms could enable more clinically relevant tools. Likewise, statistically powered studies, addressing questions such as testing turnaround times, accuracy, reproducibility and intra-observer variability, data quality used to train these algorithms, including the technical quality of scanners is critically important. Before clinical adoption of IA algorithms, guidelines for approval by regulatory agencies still must be defined, as well the paths for payer reimbursement. Most importantly, there needs to be strong consideration regarding how IA algorithms should be used within the clinical workflow, particularly with regards to current concerns about data privacy and cyber security. Missing from all of these approaches, except for the OptraScan platform, is a robust, secure data and image management system as part of the algorithm's software that both secures the patient's personal and diagnostic data and integrates this information into a health system's electronic medical record IT infrastructure. Regardless, the potential for computer IA algorithms to aid pathologists by improving diagnostic accuracy of PD-L1 scoring and to provide additional information beyond visual interpretation could transform the practice of pathology and improve patient outcomes.

FUNDING

None declared.

DISCLOSURE

Drs Inge and Dennis are employed by Roche Tissue Diagnostics.

REFERENCES

1. Chen L, Han X. Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future. *J Clin Invest*. 2015;125(9):3384–3391.
2. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012;366(26):2443–2454.
3. Ferris RL, Blumenschein Jr G, Fayette J, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med*. 2016;375(19):1856–1867.
4. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823–1833.
5. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet*. 2016;387(10031):1909–1920.
6. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133–150.
7. Wang DY, Salem JE, Cohen JV, et al. Fatal toxic effects associated with immune checkpoint inhibitors: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4(12):1721–1728.
8. Bajwa R, Cheema A, Khan T, et al. Adverse effects of immune checkpoint inhibitors (programmed death-1 inhibitors and cytotoxic T-lymphocyte-associated protein-4 inhibitors): results of a retrospective study. *J Clin Med Res*. 2019;11(4):225–236.
9. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther*. 2015;14(4):847–856.
10. Kerr KM, Nicolson MC. Non-small cell lung cancer, PD-L1, and the pathologist. *Arch Pathol Lab Med*. 2016;140(3):249–254.
11. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014;515(7528):563–567.
12. Hellmann MD, Nathanson T, Rizvi H, et al. Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell*. 2018;33(5):843–852.e4.
13. Humphries MP, McQuaid S, Craig SG, et al. Critical appraisal of programmed death ligand 1 reflex diagnostic testing: current standards and future opportunities. *J Thorac Oncol*. 2019;14(1):45–53.
14. Brunnstrom H, Johansson A, Westbom-Fremer S, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. *Mod Pathol*. 2017;30(10):1411–1421.
15. Cooper WA, Russell PA, Cherian M, et al. Intra- and interobserver reproducibility assessment of PD-L1 biomarker in non-small cell lung cancer. *Clin Cancer Res*. 2017;23(16):4569–4577.
16. Tsao MS, Kerr KM, Kockx M, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of Blueprint Phase 2 Project. *J Thorac Oncol*. 2018;13(9):1302–1311.

17. Wang C, Hahn E, Slodkowska E, et al. Reproducibility of PD-L1 immunohistochemistry interpretation across various types of genitourinary and head/neck carcinomas, antibody clones, and tissue types. *Hum Pathol*. 2018;82:131—139.
18. Downes MR, Slodkowska E, Katabi N, Jungbluth AA, Xu B. Interobserver and intraobserver agreement of PD-L1 scoring in head and neck squamous cell carcinoma (HSCC), urothelial carcinoma (UC), and breast carcinoma (BC). *Histopathology*. 2019;76(2):191—200.
19. Reisenbichler ES, Han G, Bellizzi A, et al. Prospective multi-institutional evaluation of pathologist assessment of PD-L1 assays for patient selection in triple negative breast cancer. *Mod Pathol*. 2020. https://doi.org/10.1038/s41379-020-0544-x.
20. Schwamborn K, Ammann JU, Knuchel R, et al. Multicentric analytical comparability study of programmed death-ligand 1 expression on tumor-infiltrating immune cells and tumor cells in urothelial bladder cancer using four clinically developed immunohistochemistry assays. *Virchows Arch*. 2019;475:599—608.
21. Sogani J, Allen Jr B, Dreyer K, McGinty G. Artificial intelligence in radiology: the ecosystem essential to improving patient care. *Clin Imaging*. 2019;59(1):A3—A6.
22. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167—175.
23. Zarella MD, Bowman D, Aeffner F, et al. A practical guide to whole slide imaging: a white paper from the Digital Pathology Association. *Arch Pathol Lab Med*. 2019;143(2):222—234.
24. Aeffner F, Zarella MD, Buchbinder N, et al. Introduction to digital image analysis in whole-slide imaging: a white paper from the Digital Pathology Association. *J Pathol Inform*. 2019;10:9.
25. Lange H. Digital pathology: a regulatory overview. *Lab Med*. 2011;42(10):587—591.
26. Barnes M, Srinivas C, Bai I, et al. Whole tumor section quantitative image analysis maximizes between-pathologists' reproducibility for clinical immunohistochemistry-based biomarkers. *Lab Invest*. 2017;97(12):1508—1515.
27. Samuel AL. Some studies in machine learning using the game of checkers (reprinted from Journal of Research and Development, Vol. 3, 1959) *IBM J Res Develop*. 2000;44(1-2):207—226.
28. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. 2019;16:703—715.
29. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, CM-Path AI in Histopathology Working Group. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol*. 2019;249(2):143—150.
30. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253—e261.
31. Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology image analysis using segmentation deep learning algorithms. *Am J Pathol*. 2019;189(9):1686—1698.
32. Reis S, Gazinska P, Hipwell JH, et al. Automated classification of breast cancer stroma maturity from histological images. *IEEE Trans Biomed Eng*. 2017;64(10):2344—2352.
33. Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol*. 2018;31(10):1502—1512.
34. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199—2210.
35. Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. 2016;7:12474.
36. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25(7):1054—1056.
37. McLaughlin J, Han G, Schalper KA, et al. Quantitative assessment of the heterogeneity of PD-L1 expression in non-small-cell lung cancer. *JAMA Oncol*. 2016;2(1):46—54.
38. Falk AT, Yazbeck N, Guibert N, et al. Effect of mutant variants of the KRAS gene on PD-L1 expression and on the immune microenvironment and association with clinical outcome in lung adenocarcinoma patients. *Lung Cancer*. 2018;121:70—75.
39. Parra ER, Villalobos P, Behrens C, et al. Effect of neoadjuvant chemotherapy on the immune microenvironment in non-small cell lung carcinomas as determined by multiplex immunofluorescence and image analysis approaches. *J Immunother Cancer*. 2018;6(1):48.
40. Parra ER, Villalobos P, Zhang J, et al. Immunohistochemical and image analysis-based study shows that several immune checkpoints are co-expressed in non-small cell lung carcinoma tumors. *J Thorac Oncol*. 2018;13(6):779—791.
41. Saylor J, Ma Z, Goodridge HS, et al. Spatial mapping of myeloid cells and macrophages by multiplexed tissue staining. *Front Immunol*. 2018;9:2925.
42. Koelzer VH, Gisler A, Hanhart JC, et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology*. 2018;73(3):397—406.
43. Kim HN, Jang J, Heo YJ, et al. PD-L1 expression in gastric cancer determined by digital image analyses: pitfalls and correlation with pathologist interpretation. *Virchows Arch*. 2020;476:243—250.
44. Humphries MP, Hynes S, Bingham V, 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.
45. Kapil A, Meier A, Zuraw A, et al. Deep semi supervised generative learning for automated tumor proportion scoring on NSCLC tissue needle biopsies. *Sci Rep*. 2018;8(1):17343.
46. Taylor CR, Jadhav AP, Gholap A, et al. A multi-institutional study to evaluate automated whole slide scoring of immunohistochemistry for assessment of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Appl Immunohistochem Mol Morphol*. 2019;27(4):263—269.
47. Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. *JAMA Oncol*. 2017;3(8):1051—1058.
48. Rimm DL, Han G, Taube JM, et al. Reanalysis of the NCCN PD-L1 companion diagnostic assay study for lung cancer in the context of PD-L1 expression findings in triple-negative breast cancer. *Breast Cancer Res*. 2019;21(1):72.
49. Ilie M, Hofman V, Dietel M, Soria JC, Hofman P. Assessment of the PD-L1 status by immunohistochemistry: challenges and perspectives for therapeutic strategies in lung cancer patients. *Virchows Arch*. 2016;468(5):511—525.
50. Lu S, Stein JE, Rimm DL, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. *JAMA Oncol*. 2019;5(8):1195—1204.
51. Carstens JL, Correa de Sampaio P, Yang D, et al. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun*. 2017;8:15095.
52. Gorris MAJ, Halilovic A, Rabold K, et al. Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. *J Immunol*. 2018;200(1):347—354.
53. Lazarus J, Maj T, Smith JJ, et al. Spatial and phenotypic immune profiling of metastatic colon cancer. *JCI Insight*. 2018;3(22). Available at: https://insight.jci.org/articles/view/121932. Accessed June 8, 2020.
54. Zhang W, Hubbard A, Jones T, et al. Fully automated 5-plex fluorescent immunohistochemistry with tyramide signal amplification and same species antibodies. *Lab Invest*. 2017;97(7):873—885.
55. Lim JCT, Yeong JPS, Lim CJ, et al. An automated staining protocol for seven-colour immunofluorescence of human tissue sections

for diagnostic and prognostic use. *Pathology*. 2018;50(3):333–341.

56. Surace M, DaCosta K, Huntley A, et al. Automated multiplex immunofluorescence panel for immuno-oncology studies on formalin-fixed carcinoma tissue specimens. *J Vis Exp*. 2019;143. Available at: https://www.jove.com/video/58390/automated-multiplex-immunofluorescence-panel-for-immuno-oncology. Accessed June 8, 2020.
57. Althammer S, Tan TH, Spitzmuller A, et al. Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. *J Immunother Cancer*. 2019;7(1):121.
58. Silva MA, Ryall KA, Wilm C, Caldara J, Grote HJ, Patterson-Kane JC. PD-L1 immunostaining scoring for non-small cell lung cancer based on immunosurveillance parameters. *PLoS One*. 2018;13(6):e0196464.