USCAP

ARTICLE

Comparison of continuous measures across diagnostic PD-L1 assays in non-small cell lung cancer using automated image analysis

Moritz Widmaier[1] · Tobias Wiestler[1] · Jill Walker[2] · Craig Barker[3] · Marietta L. Scott[3] · Farzad Sekhavati[1] · Alexei Budco[1] · Katrin Schneider[1] · Felix J. Segerer[1] · Keith Steele[4] · Marlon C. Rebelatto[4]

Abstract
Tumor programmed cell death ligand-1 (PD-L1) expression is a key biomarker to identify patients with non-small cell lung cancer who may have an enhanced response to anti-programmed cell death-1 (PD-1)/PD-L1 treatment. Such treatments are used in conjunction with PD-L1 diagnostic immunohistochemistry assays. We developed a computer-aided automated image analysis with customized PD-L1 scoring algorithm that was evaluated via correlation with manual pathologist scores and used to determine comparability across PD-L1 immunohistochemistry assays. The image analysis scoring algorithm was developed to quantify the percentage of PD-L1 positive tumor cells on scans of whole-slide images of archival tumor samples from commercially available non-small cell lung cancer cases, stained with four immunohistochemistry PD-L1 assays (Ventana SP263 and SP142 and Dako 22C3 and 28-8). The scans were co-registered and tumor and exclusion annotations aligned to ensure that analysis of each case was restricted to comparable tissue areas. Reference pathologist scores were available from previous studies. F1, a statistical measure of precision and recall, and overall percentage agreement scores were used to assess concordance between pathologist and image analysis scores and between immunohistochemistry assays. In total, 471 PD-L1-evaluable samples were amenable to image analysis scoring. Image analysis and pathologist scores were highly concordant, with F1 scores ranging from 0.8 to 0.9 across varying matched PD-L1 cutoffs. Based on F1 and overall percentage agreement scores (both manual and image analysis scoring), the Ventana SP263 and Dako 28-8 and 22C3 assays were concordant across a broad range of cutoffs; however, the Ventana SP142 assay showed very different characteristics. In summary, a novel automated image analysis scoring algorithm was developed that was highly correlated with pathologist scores. The algorithm permitted quantitative comparison of existing PD-L1 diagnostic assays, confirming previous findings that indicate a high concordance between the Ventana SP263 and Dako 22C3 and 28-8 PD-L1 immunohistochemistry assays.

These authors contributed equally: Moritz Widmaier, Tobias Wiestler

Supplementary information The online version of this article (https://doi.org/10.1038/s41379-019-0349-y) contains supplementary material, which is available to authorized users.

✉ Moritz Widmaier
   mwidmaier@definiens.com

1   Definiens AG, Munich, Germany

2   Oncology Companion Diagnostics Unit, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK

3   Diagnostic Development Unit, Precision Medicine, R&D Oncology, AstraZeneca, Cambridge, UK

4   MedImmune, Gaithersburg, MD, USA

Introduction

Tumors can evade the immune system via exploitation of inhibitory checkpoint pathways that suppress antitumor T-cell responses [1]. In the programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1) pathway, the PD-L1 expressed by tumor or tumor-infiltrating immune cells binds to PD-1, inhibiting T-cell receptor signaling and blocking antitumor immune response [2–4]. Antibodies targeting PD-1 or PD-L1 can block this interaction, thus resuming antitumor response [2].

The introduction of immune checkpoint inhibitors has transformed the treatment landscape for several cancers, including non-small cell lung cancer. Tumor PD-L1 expression is a key biomarker to identify patients who may have an enhanced response to non-small cell lung

cancer treatment using PD-1 or PD-L1 inhibitors [5–10]. Each treatment is currently used in conjunction with a Food and Drug Administration-approved individual diagnostic immunohistochemistry assay, as either a companion or complementary diagnostic test, to assess PD-L1 expression levels on malignant tumor and/or immune cells. Pembrolizumab, for example, is approved for use in metastatic non-small cell lung cancer as first-line treatment in patients with ≥ 50% of tumor cells expressing PD-L1, as determined by a Food and Drug Administration-approved test [11], such as the Dako IHC PD-L1 22C3 pharmDx companion assay used in the trial underlying approval [11, 12]. Nivolumab is approved for use in patients with metastatic non-small cell lung cancer who have progressed on/after platinum-based chemotherapy [13], with Dako PD-L1 28-8 pharmDx approved as a complementary PD-L1 diagnostic test [14]. Other commercially available PD-L1 immunohistochemistry tests include the Ventana SP142 and SP263 assays [15–17]. SP263, for example, is approved as a companion diagnostic test with durvalumab, which was recently approved in the European Union for use in locally advanced, unresectable non-small cell lung cancer in adults whose tumors express PD-L1 on ≥ 1% of tumor cells and whose disease has not progressed following platinum-based chemoradiation therapy [18]. All four assays have been developed independently with different antibody clones, immunohistochemistry protocols, scoring algorithms, and cutoffs to define high versus low PD-L1 expression levels [19–22]. However, previous studies using commercial non-small cell lung cancer tumor samples have shown strong concordance for PD-L1 tumor cell staining at different cutoffs for three of the four immunohistochemistry assays (i.e., the Ventana SP263 and Dako 22C3 and 28-8 assays), with less agreement using the Ventana SP142 assay [23–25].

Recent studies have shown that the scoring variation between different pathologists can be an intrinsic source of error [25–27]. For example, a recent study noted that the variability between different pathologists who were scoring the same stained samples appeared higher than the variability between different immunohistochemistry assays scored by a single reader, as reflected in the overall percentage agreement scores [25]. Thus, the practice of assigning a pathologist score manually may lead to inconsistent results. Automated image analysis may provide an aided scoring tool for pathologists to reduce inter- and intra-reader variability and increase scoring throughput (e.g., via a time advantage by eliminating the need for manual area selection on stained samples).

Here we report an extension of previous studies [24, 25] using archival tumor samples from commercially available non-small cell lung cancer cases in which automated image analysis with a customized PD-L1 scoring system was developed, evaluated via correlation with manual pathologist scores, and then used to determine comparability across the four PD-L1 immunohistochemistry assays, based on a more quantitative comparison. The PD-L1 scoring used in this analysis was developed to detect stained tumor cells, owing to their use in commercially available companion tests [24, 25].

Methods

Tumor samples and assays

As reported previously [25], 500 archival tumor resection samples (formalin-fixed, paraffin-embedded blocks) were obtained from commercially available non-small cell lung cancer cases (Asterand; ProteoGenex; Tissue Solutions). Of these, 493 were evaluable for PD-L1 expression using the Ventana SP263 and Dako 28-8 and 22C3 assays and read in batches on an assay-by-assay basis by a single manufacturer-trained pathologist in a Clinical Laboratory Improvement Amendments program-certified laboratory (Hematogenix; Tinley Park, IL, USA) (Fig. 1). Using
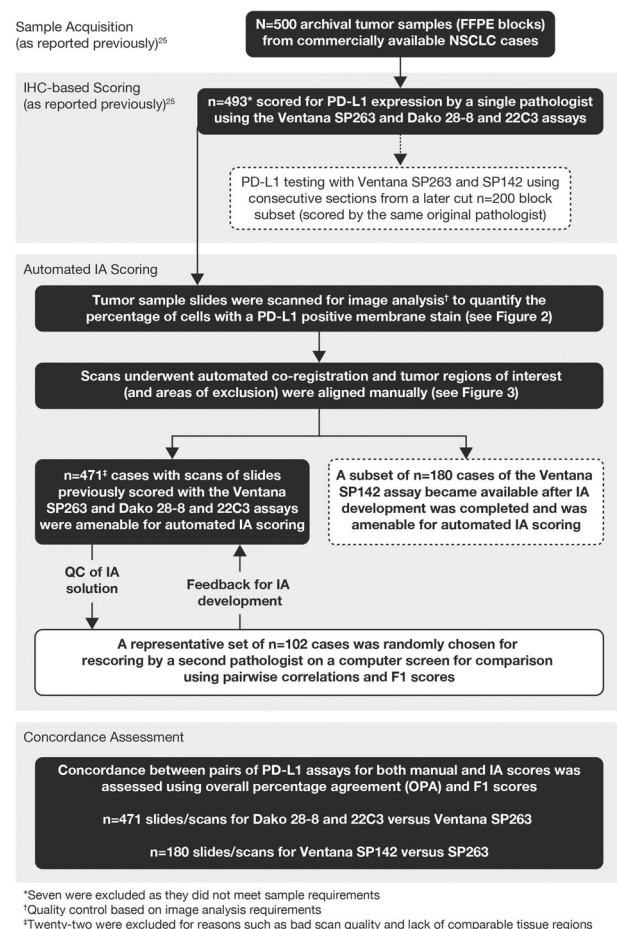

Fig. 1 Study design with IA scoring algorithm

*Seven were excluded as they did not meet sample requirements
†Quality control based on image analysis requirements
‡Twenty-two were excluded for reasons such as bad scan quality and lack of comparable tissue regions

consecutive sections from a later cut ($n = 200$) block subset, PD-L1-evaluable samples were scored by the same original pathologist using the Ventana SP263 and SP142 assays (Fig. 1). Within-block concordance for repeated SP263 staining has been previously shown [28].

Image analysis scoring algorithm

An image analysis scoring algorithm (Fig. 1) [29] was developed (based on the scores for 70 non-small cell lung cancer cases) to quantify the percentage of tumor (neoplastic) cells (excluding identified immune cells) with a positive membrane stain on a full-slide scanned image ("scan") of all samples for the four PD-L1 immunohistochemistry assays (Ventana SP263, Dako 28-8, Dako 22C3, and Ventana SP142) (Figs. 1 and 2). Slides were scanned on an Aperio Scanscope scanner at 20x magnification. The scans were automatically co-registered [30] and tumor as well as exclusion annotations were aligned manually to ensure that analysis of each case was restricted to tissue areas comparable across the four assays (Figs. 1 and 3). Pathologist scores for the

percentage of membrane-positive tumor cells, generated via microscope ("slides") for the original physical samples, were available from previous studies [24, 25] and used as the reference for the immunohistochemistry assay comparisons. The areas of pathologist assessment were not annotated, but the areas analyzed comprise most of the tumor region and hence overlap with the scanned and image analysis-analyzed regions. For evaluation of the quality of the automatically generated image analysis scores, a randomly chosen subset of the scans (using the Ventana SP263 and Dako 28-8 and 22C3 assays) were rescored by a second pathologist on a computer screen (Fig. 1). In addition to being blinded to the immunohistochemistry assay, the second pathologist only used comparable regions for scoring.

Overall percentage agreement

Total agreement for positive and negative ratings between two immunohistochemistry assays was measured as overall percentage agreement [31], taking into account concordant classification of samples above and below the thresholds
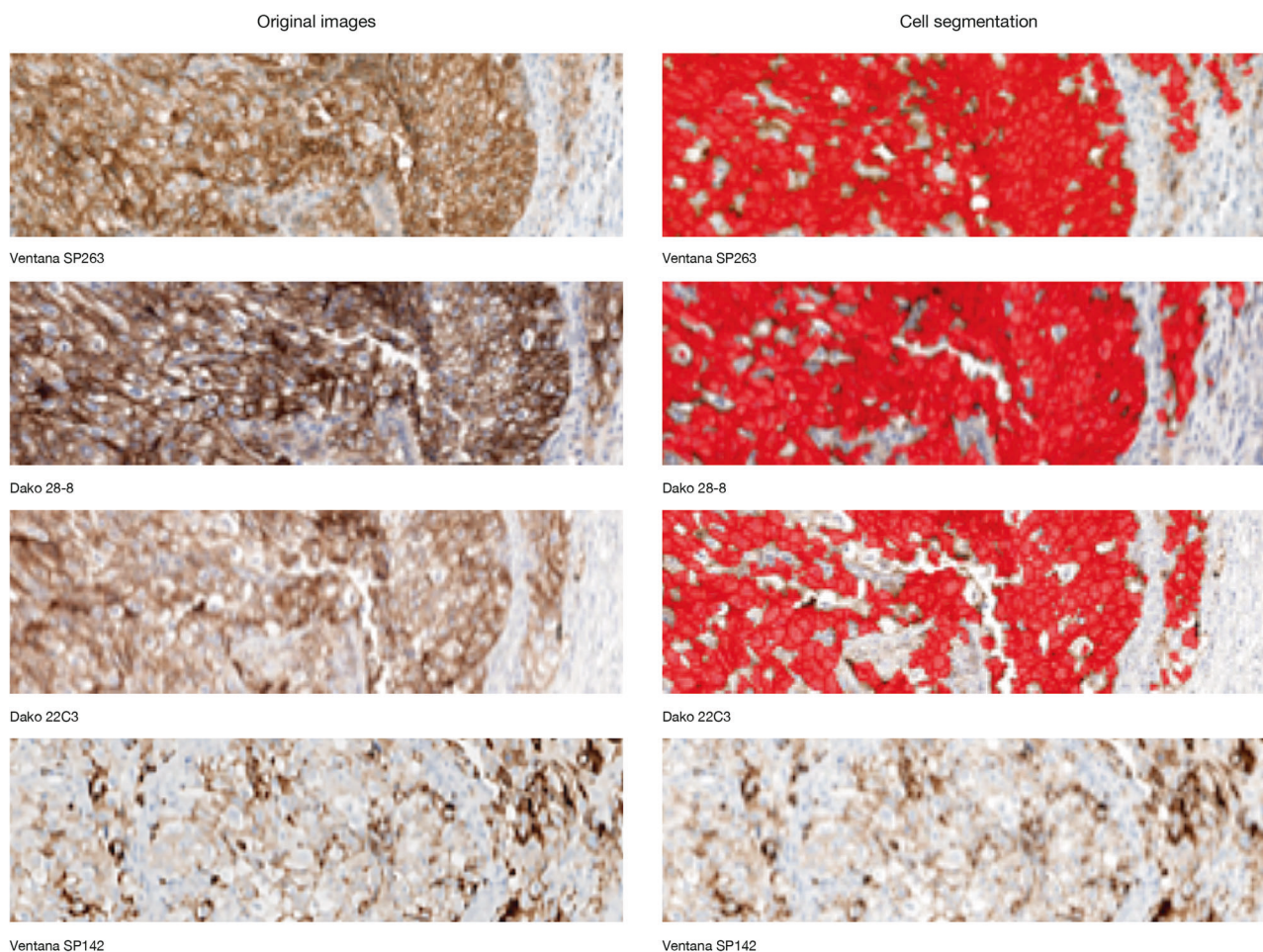

Fig. 2 Segmentation of positive PD-L1 cells. Due to differences in intensity and presentation of stain on tumor cell membranes, automated detection of positive cells using the same (assay-independent) thresholds revealed differences between the assays
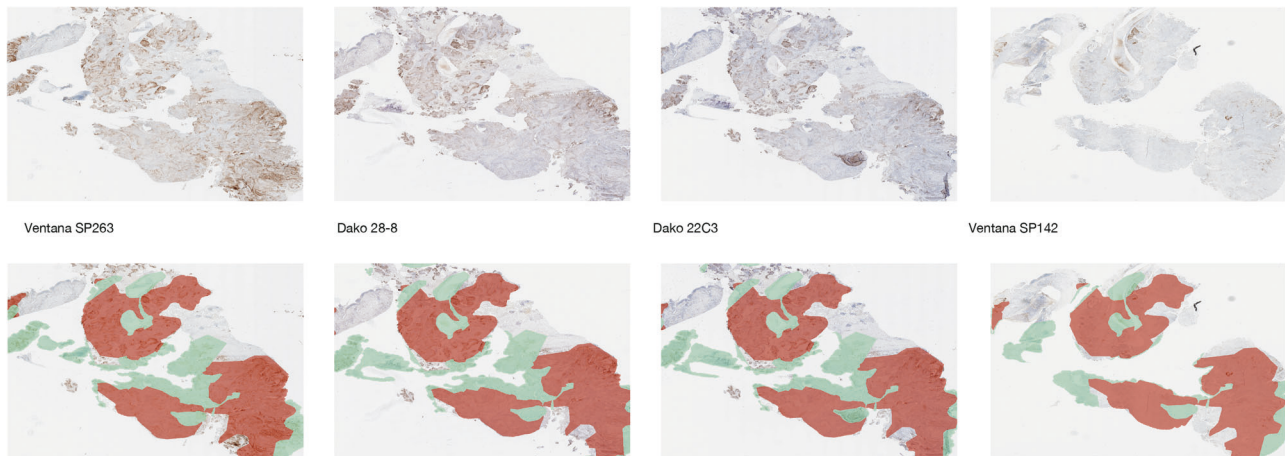
Fig. 3 Co-registered regions of interest. Top panel: original scanned images. Bottom panel: manual exclusion of artefacts on each scan (scanner artefacts and staining artefacts) and co-registered annotation transfer (tumor center, exclusion) [30]

applied for each assay:

$$OPA = \frac{TP + TN}{All\ samples}$$

in which TP and TN denote "True" positive and negative, respectively.

F1 score

The F1 score was used to measure concordance for "positive" ratings (i.e., samples above a given cutoff) between different immunohistochemistry assays or between pathologist and image analysis scores from the same immunohistochemistry assay at any given cutoff pair from 1 to 99% positive cells (in 1% increments) [32]:

$$F1\ score = \frac{2^*TP}{2^*TP + FN + FP}$$

in which TP denotes "True" positive (i.e., both assays agree), and FP or FN denote "False" positive or negative (i.e., the classifications based on the assays disagreeing at this cutoff pair).

For comparisons of image analysis scores against pathologist scores, the pathologist ratings can be considered as "ground truth"; thus, concordant ratings based on image analysis are "True" positive and discordant ratings are "False" positive and "False" negative. Importantly, although for comparisons between pathologists either can be considered as "ground truth", the calculation gives the same result. Following validation of the image analysis scoring algorithm, F1 was used to compare immunohistochemistry assays based on both pathologist and image analysis scores.

Unlike overall percentage agreement, which provides an intuitive score of the overall agreement in ratings, the F1 score focuses on agreement for positive patients which, in some instances, can be more important (e.g., for patient segmentation). For this analysis, both F1 and overall percentage agreement were used since they provide different types of information.

Image analysis optimization

As an ad-hoc analysis, a new image analysis workflow was introduced that allowed dynamic (re-)assessment of tumor cell positivity by varying different thresholds. In brief, the workflow can be outlined as follows: tumor cells were segmented in a scan (regardless of the presence of stain), with its features extracted and thresholds for these features optimized to maximize correlation with the Ventana SP263 assay reference data from previous studies [24, 25] (see Section S1 in the Supplementary Methods for more details). To test the robustness of this approach and to avoid overfitting, data from the complete dataset were split into two groups: one for training and another for testing.

Results

Patients and tumor samples

As reported previously, PD-L1-evaluable tumor (neoplastic) resection samples were obtained from 493 patients with Stage I–IV non-small cell lung cancer, the majority of whom (75.3%) were Caucasian and approximately half each had squamous and non-squamous histology [25]. Of the 493 slides previously rated by pathologists, 471 were amenable for automated image analysis scoring using the Ventana SP263 and Dako 28-8 and 22C3 immunohistochemistry assays (reasons for exclusion included poor scan quality and lack of comparable tissue regions) (Fig. 1). Of the $n = 200$ block subset previously scored using the Ventana SP142 assay, 180 were evaluable for image analysis (Fig. 1).

Comparison between pathologist scores and image analysis scores

Using a randomly chosen subset of scans ($n = 102$) for which a second pathologist was blinded to the assays (but based on the assumption that the second pathologist's ratings were "ground truth"), the automated image analysis achieved high to very high positive linear correlation [33] with the

pathologist scores generated on the same scans and regions (e.g., Spearman and Pearson correlation coefficients were in the range of 0.83–0.88 and 0.94–0.95, respectively; Fig. 4a). The high correlation was reflected by high F1 concordance values, which center around 1:1 matched cutoff pairs (Fig. 4b). Closer examination of 1:1 matched assay cutoffs revealed reasonable F1 concordance (0.8–0.9) for the Ventana SP263 and Dako 28-8 immunohistochemistry assays for
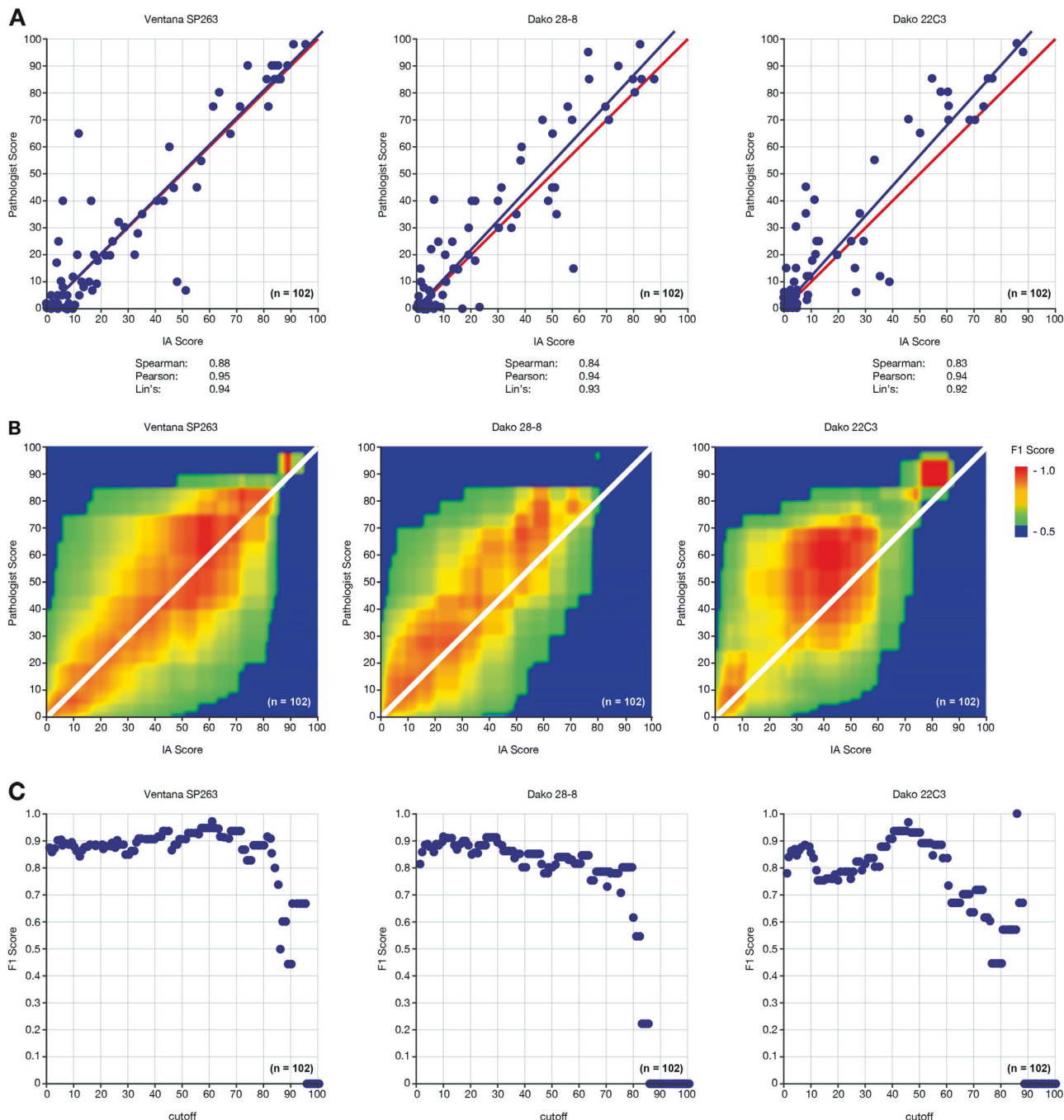
Fig. 4 a Pairwise correlations, **b** F1 concordance scores, and **c** F1 scores from 1:1 matched assay cutoffs for pathologist versus IA scores. A randomly chosen subset of the scans ($n = 102$) was rescored by a second pathologist on a computer screen (blinded to the assay) using only comparable regions for scoring. In panel **a**, the blue lines denote "best fit" or linear regression lines

cutoffs of 3% up to 65% (Fig. 4c). For Dako 22C3, a similar but slightly inferior profile was observed, with comparably low concordance around the 20% threshold. Generally, for lower (and even more so for higher) cutoff pairs, concordance was reduced. The latter is at least partially explained by the low number of strongly positive cases and slightly lower sensitivity of the automated image analysis (see the linear regression fit line in Fig. 4a). Similar results, with high linear correlation and F1 concordance, were observed when comparing the automated image analysis with mean readout scores from the 2 pathologists (Fig. S1); high correlation was also observed between the image analysis score and each pathologist independently.

Overall percentage agreement between assays

In the overall percentage agreement calculations, the manual scores from previous publications based on slides [24, 25] (Fig. 5a) and the newly produced image analysis scores based on scans (Fig. 5b) displayed very similar results, further confirming the previous results and,

indirectly, the quality of the image analysis results. Consistently, the automated image analysis scores displayed similar results to the mean readout scores from the 2 pathologists (Fig. S2).

F1 scores between assays

In this dataset, the numbers of PD-L1 negative (below threshold) samples exceeded the number of PD-L1 positive samples at most cutoffs; thus, overall percentage agreement scores were largely driven by the concordance of negative samples. F1 scores, however, restrict the comparisons to the more relevant, albeit smaller, group of positive samples. Both overall percentage agreement and F1 scoring showed concordance between Ventana SP263, Dako 28-8, and Dako 22C3 immunohistochemistry assays; however, Ventana SP142 showed very different characteristics (Fig. 6). Small differences between the previous manual scores (Fig. 6a) and automated scores (Fig. 6b) were observed. Compared with the human pathologist, the automated image analysis of the Dako 22C3 immunohistochemistry assay
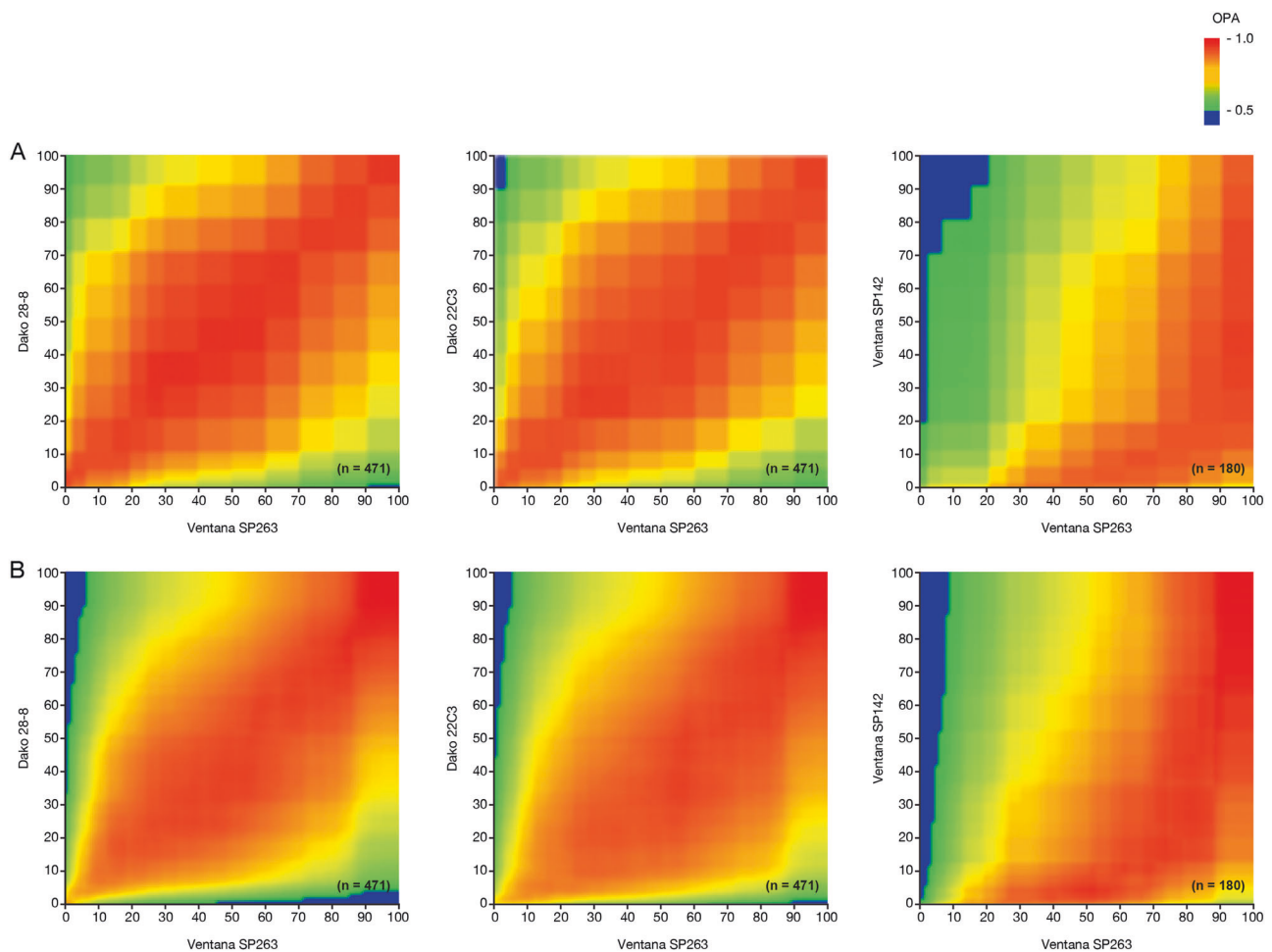

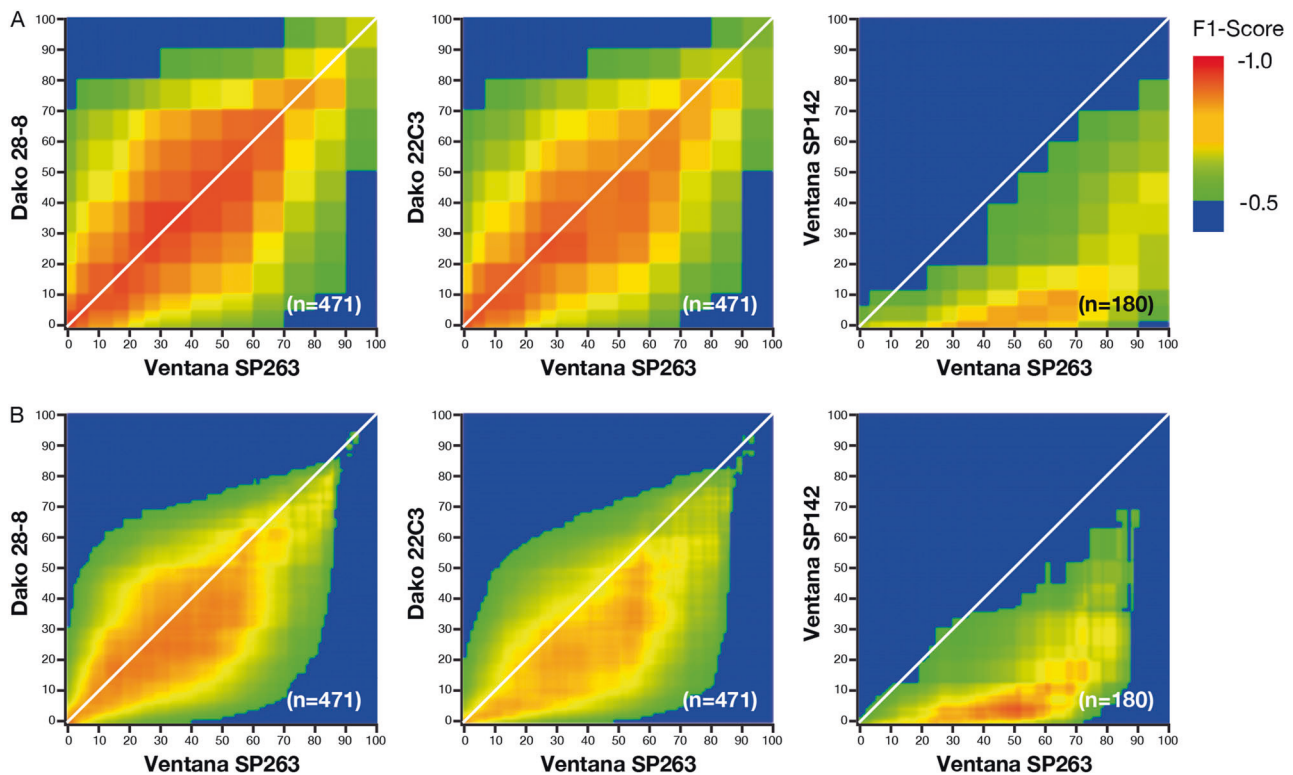Fig. 5 OPA scores between pairs of assays for **a** pathologist scores [24, 25] and **b** IA scores

Fig. 6 F1 scores between pairs of assays for **a** pathologist scores [25] and **b** IA scores

identified slightly lower percentages of positive cells, as evident in the slight skew in the F1 plot (Fig. 6b).

Image analysis optimization

As an ad-hoc analysis, a new image analysis workflow was introduced (see Section S1 in the Supplementary Methods) that allowed dynamic (re-)assessment of tumor cell positivity by varying different thresholds (Fig. 7a, b). An optimal set of thresholds was identified by iterating through thresholds for Dako 22C3 and comparing the results against the established image analysis scores from Ventana SP263 (Fig. 7c). Using the complete dataset ($n = 470$), split into training and testing groups, two immunohistochemistry-related features and their thresholds were identified that yielded the best results: a ratio membrane to overall diaminobenzidine intensity of >1.02 and a mean overall diaminobenzidine intensity in the membrane of >17.72. A linear correlation between both assays (initial [Fig. 7d] and after optimization [Fig. 7e]) was reached that eliminated the skewness from the F1 plot (Fig. 7f).

Discussion

We showed that a novel automated image analysis scoring algorithm can be used to determine tumor cell PD-L1

positivity in patients with non-small cell lung cancer and that it demonstrates high analytical concordance with pathologist ratings, permitting quantitative comparison of existing immunohistochemistry assays in order to confirm previous findings [25]. Both the manual and automated image analysis approaches showed that the Ventana SP263, Dako 28-8, and Dako 22C3 assays are highly concordant for a broad range of cutoffs on an analytical level, as reflected in both overall percentage agreement and F1 scoring, while the Ventana SP142 assay showed very different characteristics. The reasons for the distinct profile for SP142 may stem from different optimization of the assays, as well as a lower titration of SP142 compared to the other assays; these outcomes are consistent with numerous other studies showing divergent outcomes when comparing SP142 staining to other assays [23, 34, 35]. While the SP142 was not compared directly with the Dako 28-8 or Dako 22C3 assays, the high concordance of these two assays with SP263, in parallel with the low concordance between SP142 and SP263, suggest that SP142 would also show lower outcome concordance with the other two assays. The F1 scores highlight that the automated image analysis of Dako 22C3 identified slightly lower percentages of PD-L1 positive tumor cells than did pathologists. Importantly, differences between immunohistochemistry assays may have been surmounted by human translational capabilities in the previous study [25] (e.g., human
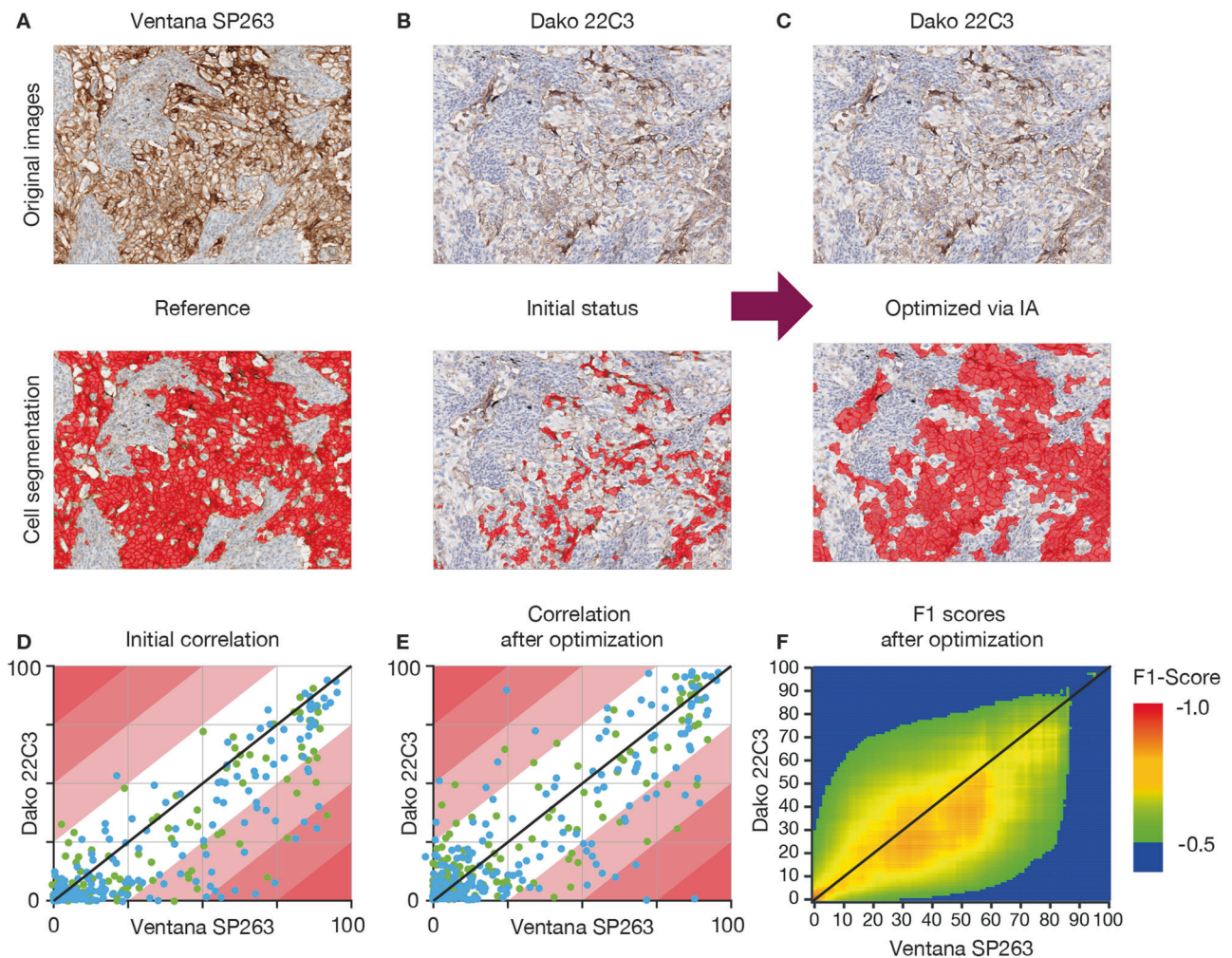
Fig. 7 IA optimization of positivity thresholds for tumor cells. In panels **d** and **e**, the points of the plotline falling on the diagonal line indicate a perfect match of the pathologist and IA scores. The red lines/regions indicate regions of increasing discordance between both ratings in 25% (absolute value) increments (e.g., the white region indicates a region where both ratings agree by ±25% and the light red region is where both ratings agree by ±50%, etc.). In panel **f** (after optimization), the skewness is removed (perhaps, however, at the cost of slightly inferior overall accuracy)

pathologists were required to "interpret" subtle differences in staining presentation between assays for which this algorithm is not tuned). In comparison to previous efforts based on pathologist ratings, no assay-specific adaptations based on interpretation were made by our algorithm (i.e., independent of the immunohistochemistry assay, each cell was tested against the same criteria for positivity). However, it was demonstrated that inter-assay differences can be minimized by optimization of image analysis morphology- and immunohistochemistry-related parameters and their thresholds, indicating that this could be used to improve consistency and concordance with pathologist ratings or other relevant measures (e.g., outcome and gene or protein expression).

Response rates to anti-PD-1 and anti-PD-L1 agents have been shown to be greater in patients whose tumors express high levels of PD-L1 compared with those expressing low or no tumor PD-L1 [1–4]. Broad access to high-quality PD-L1 testing will help clinicians to identify the most appropriate treatment option for individual patients. High concordance between most diagnostic tests has been reported earlier [23, 25]. In the recent study by Ratcliffe et al. the variability between two different pathologists scoring the same stained samples appeared higher than the variability between different assays scored by a single reader; in addition, concordance between different pathologists scoring the same slides was lower for samples with staining below 10% [25]. In the case of image analysis scoring, inter-reader variability can be reduced using digitized scoring and optimization, ensuring consistent and reproducible readouts across the board.

Our results, showing the ability of automated image analysis to assess PD-L1 status in patients with non-small cell lung cancer, complement those of a recent analysis
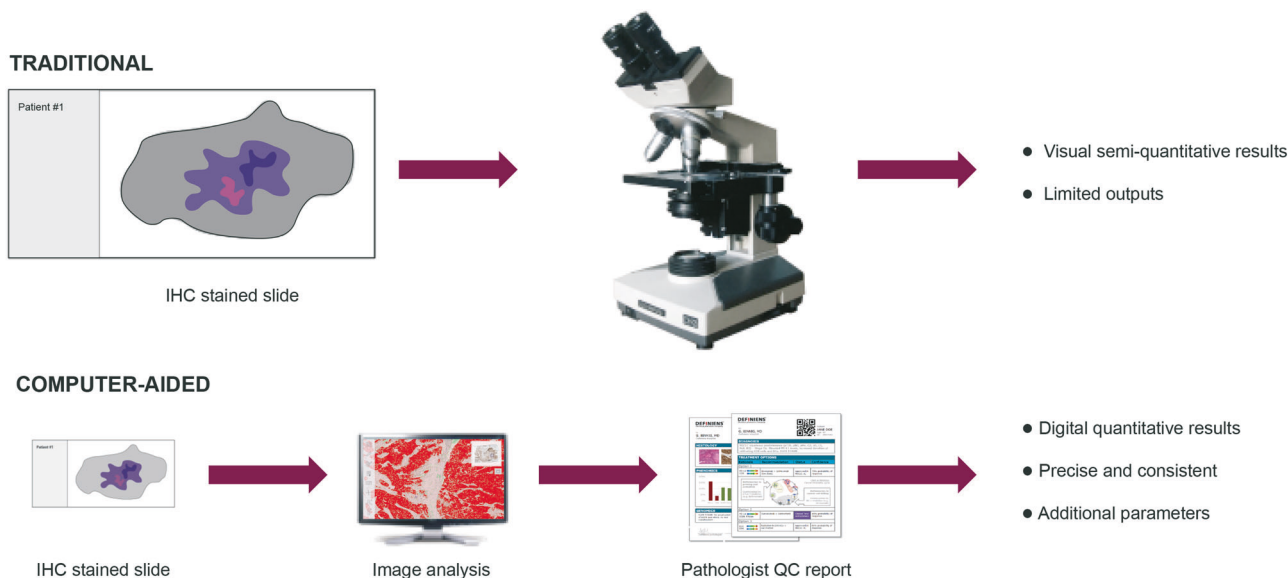
Fig. 8 Automated IA as an aided scoring tool

using data from a Phase 1/2 study, which demonstrated that an automated image analysis signature (based on combined baseline cell densities of PD-L1[+] tumor cells and CD8[+] tumor infiltrating lymphocytes) may allow better identification of responders to durvalumab monotherapy compared with manual PD-L1 scoring alone [36]. Furthermore, in a separate study that employed a deep semi-supervised and generative learning network, automated PD-L1 scoring of non-small cell lung cancer tumor tissue needle biopsies was concordant with visual scoring by pathologists [37]. It is important to note that, using a PD-L1 staining assay, pathologists occasionally fail to distinguish between tumor-infiltrating PD-L1 + macrophages from PD-L1 + tumor cells. As such, inclusion of macrophages in the image analysis-generated PD-L1 assessment may in fact lead to more reproducible outcomes. However, the concordance across the different assays (particularly Ventana SP263, Dako-22C3, and Dako 28-8), as well as between the pathologist and image analysis analyses, suggests that the impact of PD-L1 + macrophages on assay readout is minimal. As such, the scoring algorithms used in non-small cell lung cancer do not account for macrophages. Novel multiplexing assays using immunofluorescence to identify PD-L1 + macrophages are of interest and would provide greater insight into the immune-context of tumors.

A potential limitation of an image analysis approach is that scans cannot represent all of the details that are visible under a microscope (especially at higher resolution); hence, a deviation in results to a certain extent is to be expected [38]. In addition, while automated image analysis is expected to deliver results with low variance even in the low positivity range, it remains to be

demonstrated if those results are also accurate. Comparison to ratings of multiple pathologists is required to validate this further.

In summary, the digital quantitative results from automated image analysis scoring demonstrated comparable accuracy and consistency to that provided by pathologists' scoring. As such, image analysis scoring could serve as an aid, with appropriate validation, for PD-L1 diagnostic testing for pathologists in the clinical setting (Fig. 8) and help support scoring with commercial assays.

Compliance with ethical standards

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Dong H, Strome SE, Salomao DR, et al. Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. Nat Med. 2002;8:793–800.

2. Borczuk AC, Allen TC. PD-L1 and lung cancer: the era of precision-ish medicine? Arch Pathol Lab Med. 2016;140:351–4.

3. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012;12:252–64.

4. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol. 2015;33:1974–82.

5. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med. 2015;373:1627–39.

6. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med. 2015; 372:2018–28.

7. Carbognin L, Pilotto S, Milella M, et al. Differential activity of nivolumab, pembrolizumab and MPDL3280A according to the tumor expression of programmed death-ligand-1 (PD-L1): sensitivity analysis of trials in melanoma, lung and genitourinary cancers. PLoS One. 2015;10:e0130142.

8. Fehrenbacher L, Spira AI, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet. 2016;387:1837–46.

9. Gulley JL, Rajan A, Spigel DR, et al. Avelumab (MSB0010718C), an anti-PD-L1 antibody, in patients with metastatic or recurrent non-small-cell lung cancer progressing after platinum-based chemotherapy: A phase Ib trial. Paper presented at the 2015 European Cancer Congress (ECC), September 25–29, 2015, Vienna, Austria.

10. Garassino M, Vansteenkiste J, Kim J-H, et al. Durvalumab in ≥3rd-line locally advanced or metastatic, *EGFR/ALK* wild-type NSCLC: results from the phase 2 ATLANTIC Study. J Thorac Oncol. 2017;12:S10–S11. Abstr. PL04a.03

11. Merck Sharp & Dohme. Keytruda® prescribing information. Updated August 2018. Available at: http://www.merck.com/product/usa/pi_circulars/k/keytruda/keytruda_pi.pdf (Accessed November 6, 2018).

12. US Food and Drug Administration. Dako PD-L1 IHC 22C3 pharmDx. Available at: http://www.accessdata.fda.gov/cdrh_docs/pdf15/P150013c.pdf (Accessed October 4, 2018).

13. Bristol-Myers Squibb. Opdivo® prescribing information. Updated August 2018. Available at: http://packageinserts.bms.com/pi/pi_opdivo.pdf (Accessed November 6, 2018).

14. US Food and Drug Administration. Dako PD-L1 IHC 28-8 pharmDx. Available at: http://www.accessdata.fda.gov/cdrh_docs/pdf15/P150025c.pdf (Accessed October 4, 2018).

15. US Food and Drug Administration. VENTANA PD-L1 (SP142) Assay. Roche. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160002c.pdf (Accessed October 4, 2018).

16. US Food and Drug Administration. VENTANA PD-L1 (SP263) Assay. Roche. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160046C.pdf (Accessed October 4, 2018).

17. Ventana PD-L1 (SP263) Assay (CE-IVD). Product specifications. Roche. Available at: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp263-assay2.html (Accessed November 6, 2018).

18. European Medicines Agency. Imfinzi Summary of Product Characteristics. Available at: https://www.ema.europa.eu/documents/product-information/imfizi-epar-product-information_en.pdf (Accessed October 31, 2018).

19. Roach C, Zhang N, Corigliano E, et al. Development of a companion diagnostic PD-L1 immunohistochemistry assay for pembrolizumab therapy in non-small-cell lung cancer. Appl Immunohistochem Mol Morphol. 2016;24:392–7.

20. Phillips T, Simmons P, Inzunza HD, et al. Development of an automated PD-L1 immunohistochemistry (IHC) assay for non-small cell lung cancer. Appl Immunohistochem Mol Morphol. 2015;23:541–9.

21. Boyd ZS, Smith DC, Baker B, et al. Development of a PD-L1 companion diagnostic IHC assay (SP142) for atezolizumab. Cancer Immunol Res. 2016;4:B001.

22. Rebelatto MC, Midha A, Mistry A, et al. Development of a programmed cell death ligand-1 immunohistochemical assay validated for analysis of non-small cell lung cancer and head and neck squamous cell carcinoma. Diagn Pathol. 2016;11:95.

23. Büttner R, Gosney JR, Skov BJ, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. J Clin Oncol. 2017;35:3867–76.

24. Scott M, Scorer P, Lawson N, et al. Assessment of heterogeneity of PD-L1 expression in NSCLC, HNSCC and UC with Ventana SP263 assay. J Clin Oncol. 2017;35:e14503. (Abstr.)

25. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between programmed cell death ligand-1 diagnostic assays across multiple protein expression cutoffs in non-small cell lung cancer. Clin Cancer Res. 2017;23:3585–91.

26. Marchetti A, Barberis M, Franco R, et al. Multicenter comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) assays to test PD-L1 expression for NSCLC patients to be treated with immune checkpoint inhibitors. J Thorac Oncol. 2017;12:1654–63.

27. Brunnström 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;30:1411–21.

28. Scorer P, Scott M, Lawson N, et al. Consistency of tumor and immune cell programmed cell death ligand-1 expression within and between tumor blocks using the VENTANA SP263 assay. Diagn Pathol. 2018;13:47.

29. Baatz M, Zimmermann J, Blackmore CG. Automated analysis and detailed quantification of biomedical images using Definiens cognition network technology. Comb Chem High Throughput Screen. 2009;12:908–16.

30. Yigitsoy M, Schmidt G Hierarchical patch-based co-registration of differently stained histopathology slides. Proceedings of the

SPIE, Volume 10140, id. 1014009 6 pp. 2017: https://doi.org/10.1117/12.2254266.

31. US Food and Drug Administration. Guidance for industry and FDA staff. Statistical guidance on reporting results from studies evaluating diagnostic tests. Available at: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf (Accessed October 4, 2018).

32. Van Rijsbergen CJ. Information Retrieval. 2nd edn. Newton, MA: Butterworth-Heinemann; 1979.

33. Mukaka MM. Statistics Corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24:69–71.

34. Schats KA, Van Vre EA, Boeckx C, et al. Optimal evaluation of programmed death ligand-1 on tumor cells versus immune cells requires different detection methods. Arch Pathol Lab Med. 2018;142:982–91.

35. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. J Thorac Oncol. 2017;12:208–22.

36. 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;7:121.

37. 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;8:17343.

38. US Food and Drug Administration, Center for Devices and Radiological Health. Philips IntelliSite Pathology Solution (PIPS; whole slide imaging system) approval letter; April 17, 2017. Available at: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm553358.htm (Accessed October 4, 2018).