Annals of Diagnostic Pathology 50 (2021) 151590

Contents lists available at ScienceDirect

Annals of Diagnostic Pathology

journal homepage: www.elsevier.com/locate/anndiagpath

Original Contributions

Comparability of laboratory-developed and commercial PD-L1 assays in non-small cell lung carcinoma

Julia R. Naso[a], Gang Wang[b], Norbert Banyi[b], Fatemeh Derakhshan[a], Aria Shokoohi[c], Cheryl Ho[c], Chen Zhou[b], Diana N. Ionescu[b,*]

[a] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC V6T 1Z1, Canada
[b] Department of Pathology, British Columbia Cancer, Vancouver, BC V5Z 4E6, Canada
[c] Department of Medical Oncology, British Columbia Cancer, Vancouver, BC V5Z 4E6, Canada

ARTICLE INFO

Keywords:
PD-L1
Non-small cell lung carcinoma
22C3
SP263
Immunohistochemistry

ABSTRACT

PD-L1 expression in non-small cell lung cancer (NSCLC) is predictive of response to treatment with PD-1 and PD-L1 inhibitors. Different inhibitors have been developed with different PD-L1 assays, which use different PD-1 antibody clones on different immunohistochemistry platforms. Depending on instrument and reagent availability, laboratory-developed tests with cross-platform use of PD-L1 antibodies may have practical benefits over commercial assays. The 22C3 pharmDx Assay (referred to as 22C3 DAKO), the VENTANA PD-L1 SP263 Assay (referred to as SP263 VENTANA) and a lab-developed test using the 22C3 antibody on the VENTANA BenchMark ULTRA IHC/ISH system (referred to as 22C3 VENTANA) were performed on whole sections of 85 NSCLC surgical resections. All sections were independently scored by three pathologists using tumor proportion scores. Correlation coefficients for continuous scores in pairwise comparisons between assays ranged from 0.976 to 0.978. When using a 1% positivity threshold (dichotomous scores), the 22C3 DAKO assay and 22C3 VENTANA assays showed the greatest agreement (93% agreement, $\kappa = 0.86$, 95% CI 0.75–0.97), and the 22C3 DAKO and SP263 VENTANA assays tended to show slightly less agreement (84% agreement, $\kappa = 0.66$, 95% CI 0.50–0.82). When using a 50% positivity threshold (dichotomous scores), all pairwise comparisons showed similar agreement (96–99% agreement, $\kappa = 0.89$–0.97). Overall, there was no significant difference between assays at 1% or 50% thresholds ($P = .77$). These data are consistent with potential interchangeability of these assays, which may widen the scope of PD-L1 assays available to laboratories and reduce logistical barriers to testing.

1. Introduction

Immunotherapy has become part of the standard of care in non-small cell lung carcinoma (NSCLC). Available immunotherapy agents include inhibitors of PD-1 (e.g. pembrolizumab and nivolumab) and PD-L1 (e.g. atezolizumab and durvalumab). These agents re-activate anti-tumor immune responses by disrupting the immunosuppressive interaction of PD-L1 with PD-1 [1]. Immunohistochemical PD-L1 assays have shown value in predicting response to PD-1 and PD-L1 inhibitor treatment [2], though existing PD-L1 assays use a variety of different antibody clones and technical platforms [3,4].

Pembrolizumab is approved by the United States Food and Drug Administration (FDA) and European Medical Agency (EMA) for first-line treatment of advanced NSCLC with PD-L1 Tumor Proportion Score (TPS) ≥ 1%, determined using an approved companion diagnostic test [5-8]. The pharmDx 22C3 assay (Dako/Agilent, Santa Clara, CA) has FDA approval as a companion diagnostic for pembrolizumab, but can only be performed on Autostainer Link 48 instrument (Dako/Agilent, Santa Clara, CA) [5]. The VENTANA SP263 PD-L1 assay is European Conformity-In Vitro Diagnostic (CE-IVD) marked as a companion diagnostic for pembrolizumab [2], but is not presently FDA approved for this indication. The VENTANA SP263 PD-L1 assay is FDA approved as a complementary diagnostic for durvalumab treatment of locally advanced and metastatic urothelial carcinoma [1,9]. Durvalumab also has FDA approval for the treatment of unresectable stage III NSCLC that has not progressed after chemoradiation treatment [10,11]. The EMA has approved durvalumab for cases that fit the above indication and express PD-L1 in ≥1% of tumor cells [10,12].

The other PD-1 or PD-L1 inhibitors FDA-approved for treatment of advanced NSCLC are nivolumab and atezolizumab, both of which may be used clinically independent of PD-L1 expression [1,2]. However, PD-L1 testing retains predictive value for these agents and may still be used

* Corresponding author at: BC Cancer Agency, 600 W 10th Ave, Vancouver, BC V5Z 4E6, Canada.
E-mail address: dionescu@bccancer.bc.ca (D.N. Ionescu).

https://doi.org/10.1016/j.anndiagpath.2020.151590

Available online 14 August 2020

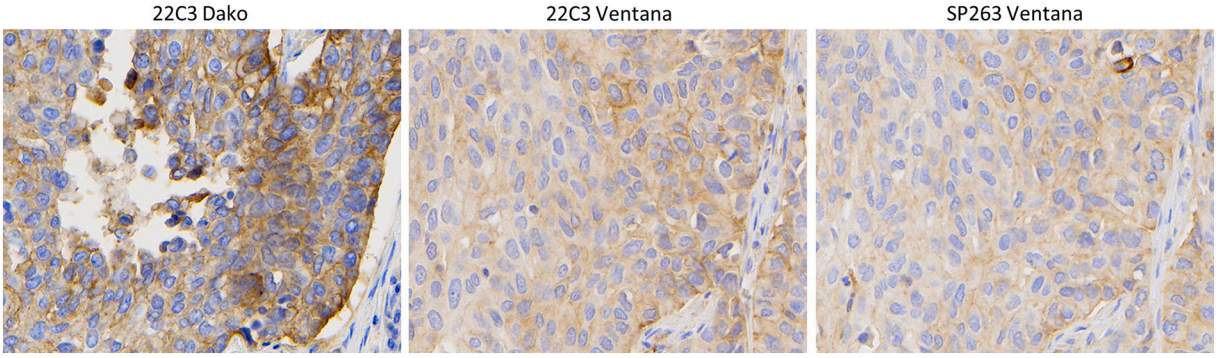

Fig. 1. Qualitative differences in PD-L1 staining between assays include more intense staining by 22C3 DAKO in some cases (X400).

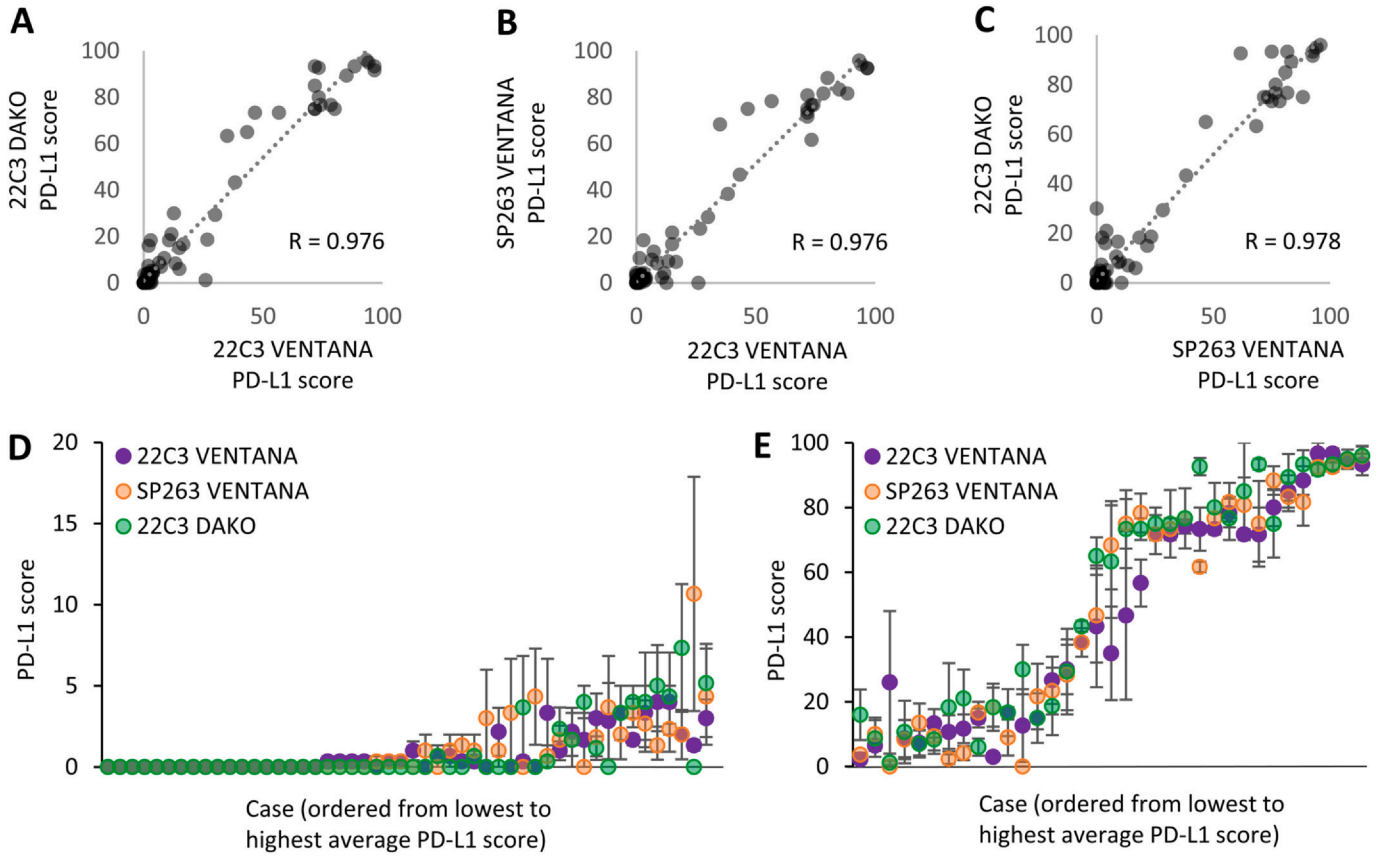

Fig. 2. Comparison of continuous PD-L1 scores for each PD-L1 assay. (A–C) show PD-L1 scores in pairwise comparisons of the assays. A linear trendline fitted to the data is shown for each plot. R values indicate Pearson correlation coefficients. (D, E) All PD-L1 scores for cases ordered by their mean PD-L1 score. Cases with an average score of ≤5% are shown in (D), and those with an average score > 5% are shown in (E). All plots show the average of the tumor proportion scores of three pathologists. Error bars indicate standard error.

in treatment decisions [2,13]. The Dako 28-8 pharmDx PD-L1 assay is FDA-cleared as a complementary diagnostic for nivolumab, and the VENTANA SP142 PD-L1 assay is FDA-cleared as a complementary diagnostic for atezolizumab [2].

Offering all the above companion and complementary PD-L1 assays is impractical for most laboratories, fueling interest in the interchangeability of assays. Laboratory-developed assays with cross-platform use of PD-L1 antibodies may also allow technology already existing in a laboratory to be used. Such assays may eliminate the need for a Dako Autostainer Link 48 for use with 22C3 PD-L1 antibody (as is required for the FDA-approved pharmDx assay), allowing greater automation than is possible on the Autostainer Link 48. However, technical limitations (such as SP263 antibody availability only in pre-diluted format), may preclude effective optimization, limiting feasible cross platform applications.

Lab developed tests (LDTs) must be properly validated in order to be considered suitable for clinical use [4]. Thus far, studies using 22C3 antibody on the VENTANA BenchMark ULTRA platform have shown promising results that this LDT may perform comparably to commercial assays [14-21]. Studies of the FDA-approved 22C3, 28–8, SP263 and SP142 assays found analytical concordance among all assays except SP142, which had lower sensitivity [13-15,22,23].

In this study, we compare PD-L1 immunohistochemistry performed using the Dako 22C3 PharmDx assay, the VENTANA SP263 assay, and a cross-platform LDT using 22C3 antibody on the VENTANA BenchMark ULTRA platform.

Table 1
Demographics of study cases.

Category	n	%
Sex		
Female	42	49%
Male	43	51%
Diagnosis		
Adenocarcinoma, non-mucinous	66	78%
Mucinous adenocarcinoma	3	4%
Adenosquamous carcinoma	1	1%
Squamous cell carcinoma	12	14%
Combined small cell carcinoma	1	1%
Non-small cell carcinoma NOS	1	1%
Sarcomatoid carcinoma	1	1%
Site		
Lung	82	96%
Bone	1	1%
Brain	2	2%
Procedure		
Lobectomy	53	62%
Wedge resection	29	34%
Other	3	4%
ALK rearrangement identified	0	0%
EGFR mutation identified	28	33%

2. Material and methods

This study was approved by the University of British Columbia Research Ethics Board under application number H18-01619. We identified surgical resections of NSCLC on which clinical PD-L1 testing was performed between March 2017 and November 2018 at BC Cancer. The clinical PD-L1 immunostained slides, referred to as the '22C3 DAKO' set, were produced using the 22C3 pharmDx Assay (Dako/Agilent, Santa Clara, CA) on an Autostainer Link 48 (Dako/Agilent, Santa Clara, CA) following the assay's instruction manual. Freshly obtained unstained slides from the block used for clinical testing were also immunostained with the following assays. All immunohistochemistry was performed at BC Cancer.

SP263 VENTANA immunostaining used the commercially developed VENTANA PD-L1 (SP263) Assay (Ventana/Roche, Tucson, AZ), on a Ventana BenchMark ULTRA IHC/ISH system (Ventana/Roche, Tucson, AZ), with an OptiView DAB IHC Detection Kit (#760-700, Ventana/Roche, Tucson, AZ), following the FDA-approved assay protocol. The 22C3 VENTANA LDT was based on a protocol from the Canadian Multicentre 22C3 IHC LDT Validation Project [21], and was performed using 22C3 PD-L1 antibody on the VENTANA BenchMark ULTRA IHC/ISH system (Ventana/Roche, Tucson, AZ), with the OptiView DAB IHC Detection Kit (#760-700, Ventana/Roche, Tucson, AZ). A cell conditioning step using ULTRA Cell Conditioning Solution (#950-224, Ventana/Roche, Tucson, AZ) was performed for 48 min, prior to a 64 min room temperature incubation with 22C3 PD-L1 antibody (#M365-3, Dako/Agilent, Santa Clara, CA) at a 1:40 dilution. Following in–house optimization, the conditions selected for use of SP263 PD-L1 antibody (#790-4905, VENTANA/ Roche, Tucson, AZ) on the Dako Omnis platform (Dako/Agilent, Santa Clara, CA) were: 30 min TRS-L pre-treatment, 30 min primary antibody treatment, 10 min rabbit linker treatment and 30 min HRP application.

All immunostained slides were independently scored by three pathologists with training and experience in PD-L1 interpretation. All sections had at least 100 viable tumor cells. Tumor proportion scores (TPS) indicate the proportion of tumor cells with partial or complete membranous PD-L1 staining and were evaluated as continues variables (0 to 100%). Scores from the three pathologists were averaged, then placed into categories of $<1\%$, 1–49% and $\geq 50\%$. Cytoplasmic staining of tumor cells and inflammatory cell staining were not scored.

Agreement across different positivity thresholds was assessed using weighted Cohen's kappa (kappa values 0.40 to 0.69 indicate weak agreement, 0.70 to 0.79 indicate moderate agreement, 0.80 to 0.89 indicate strong agreement and ≥ 0.9 indicate near perfect agreement) [13]. Inter-observer agreement was assessed using Light's kappa, which is the average Cohen's kappa among more than two raters. R values indicate Pearson correlation coefficients. F1 and overall percent agreement statistics were calculated as described previously [24]. Overall percent agreement represents the proportion of all cases with matching results in both assays. The F1 statistic is calculated in the same manner but excludes cases with negative results in both assays and double counts the number of cases with positive results in both assays. Positive percent agreement and negative percent agreement were calculated in pairwise comparisons between assays, treating the commercial assay as the gold standard, or, in comparison between commercial assays, the 22C3 assay as the gold standard. Significance testing used two-tailed Wilcoxon Signed-Rank tests for continuous scores and Chi-squared tests for categorical scores, with P-values $< .05$ considered statistically significant. Statistical analysis used RStudio version 1.2.1335.

3. Results

Eighty-five NSCLC cases from patients age 40 to 88 (median 70 years old) were identified for this study (demographics in Table 1). In-house attempts to develop an assay for SP263 on a Dako instrument did not achieve promising results (see Supplemental Fig. 1), likely due in part to the limited options for optimising a pre-diluted antibody. Comparisons between assays therefore focused on the 22C3 DAKO, SP263 VENTANA and 22C3 VENTANA assays. The 22C3 Dako assay produced qualitatively more intense staining in some samples (Fig. 1). All three assays scored on a continuous scale were highly concordant, with correlation coefficients between 0.976 and 0.978 for all pairwise comparisons between assays (Fig. 2A–C). No statistically significant differences were present in pairwise comparisons between assays for cases scoring $\leq 5\%$ in all assays (p-values between 0.62 and 0.93, Fig. 2D). However, among cases with an average score $>$ 5%, 22C3 DAKO scores were on average 5.3% higher than those in 22C3 VENTANA ($P = .0099$, Fig. 2E).

When continuous scores were converted to $<1\%$, 1–49% and $\geq 50\%$ categories, as used in clinical practice, there was no significant difference in the proportion of cases in each category between assays (Chi-squared test $P = .77$, Fig. 3A, B). If considering the 22C3 DAKO assay to be the gold standard, use of the 22C3 VENTANA assay resulted in only 6 samples (7%) being misclassified across a 1% threshold and 3 samples (4%) being misclassified across a 50% threshold (Fig. 3C). Similarly, when treating the SP263 VENTANA assay as a gold standard, the 22C3 VENTANA assay misclassified 12 samples (14%) across a 1% threshold and 2 samples (2%) samples across a 50% threshold (Fig. 3C). This rate of 'misclassification' is similar to the incidence of discordance between the two commercial assays (14 samples discordant across 1% threshold and 1 sample discordant across 50% threshold, Fig. 3C). When exploring different possible thresholds for positivity, overall percent agreement between assays increased as the threshold for positivity was increased (Fig. 3D). Percent agreement in pairwise comparisons between the assays ranged from 84% to 93% when using a 1% threshold for positivity, and from 96% to 99% when using a 50% threshold for positivity. A similar trend was found using F1 statistics (which reflect concordance specifically for positive scores without defining a gold standard) and negative percent agreement, but not for positive percent agreement (Fig. 3D).

Cohen's kappa statistics showed moderate agreement between assays when using a 1% threshold ($\kappa = 0.66$–0.86) and excellent agreement between assays when using a 50% threshold ($\kappa = 0.89$–0.97, Table 2). At the 1% threshold, the two assays using 22C3 antibody tended to show the greatest agreement with each other, whereas the assays using different antibodies and platforms (i.e. 22C3 DAKO and SP263 VENTANA) tended to show the least agreement, though

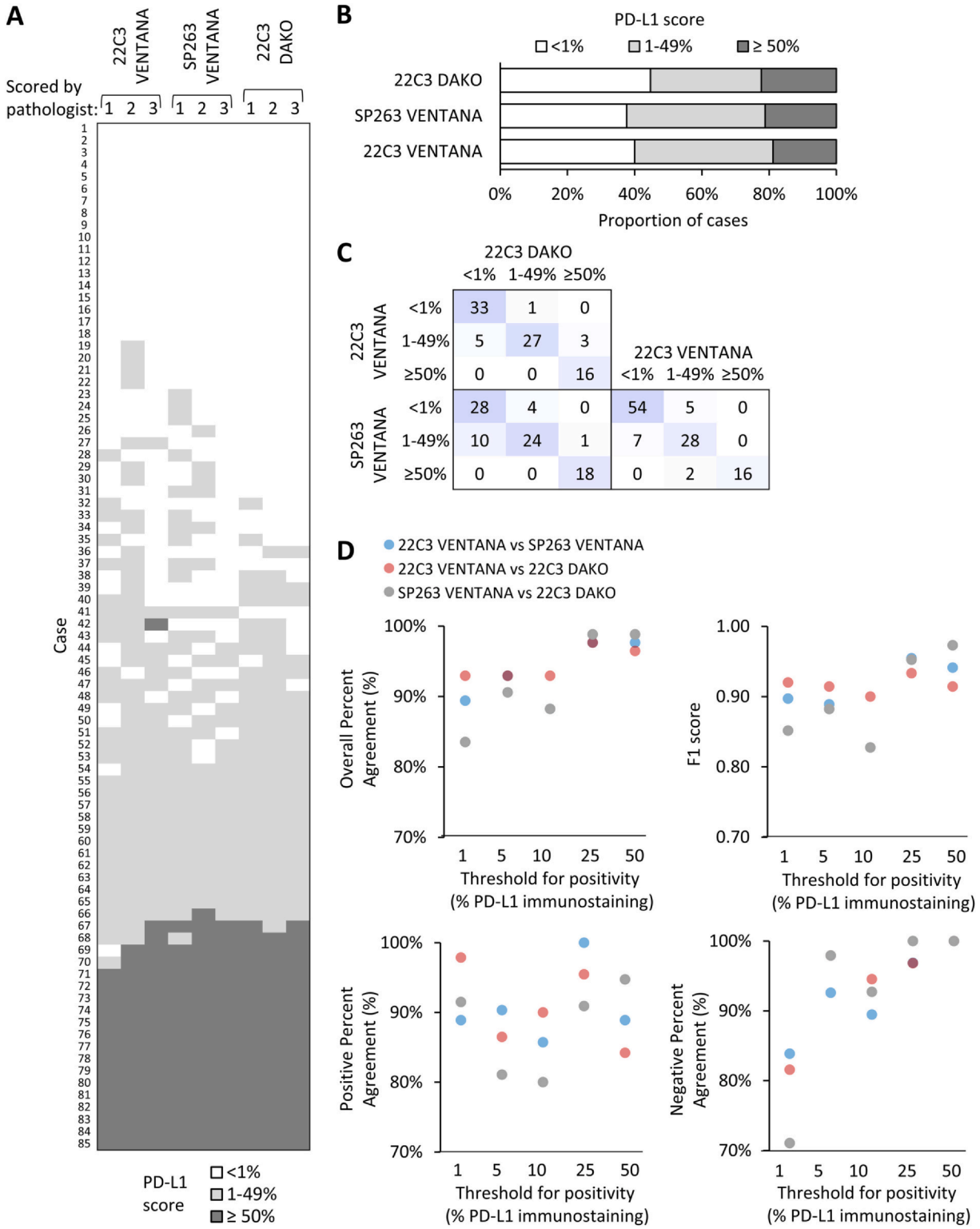

Fig. 3. Agreement between PD-L1 assays across different thresholds for positivity. (A) Individual pathologist's scores for each case, using 1% and 50% thresholds. (B) The proportion of cases scoring ≥50%, 1–49% or < 1% using each assay. (D) Overall percent agreement, F1 scores, positive percent agreement and negative percent agreement at each threshold value for PD-L1 positivity. Markers are purple where blue and pink markers overlap. Statistics were calculated using the average score of three pathologists. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

differences in κ statistics were not statistically significant.

Interobserver variability within each assay was also greater when using a 1% threshold (κ = 0.59–0.86) compared to a 50% threshold (κ = 0.88–0.98, Table 3). Using three category scoring (< 1%, 1–49% and ≥50%) interobserver agreement tended to be greatest for the 22C3 DAKO assay and lowest for the 22C3 VENTANA assay (Table 3).

4. Discussion

We find that the 22C3 DAKO, SP263 VENTANA and 22C3

Table 2
Weighted Cohen's kappa statistics for agreement between PD-L1 assays. For 'average scores' the scores of the three pathologists on a continuous scale were averaged prior to application of the indicated threshold.

Assay		1% positivity threshold		50% positivity threshold		1% and 50% thresholds (3 categories)	
		Kappa	95% CI	Kappa	95% CI	Kappa	95% CI
22C3 VENTANA vs SP263 VENTANA	Pathologist #1	0.67	0.51–0.82	0.89	0.76–1	0.81	0.71–0.90
	Pathologist #2	0.59	0.43–0.76	0.90	0.78–1	0.80	0.69–0.90
	Pathologist #3	0.91	0.88–1	0.97	0.90–1	0.96	0.89–1
	Average scores	0.7	0.55–0.86	0.93	0.83–1.0	0.85	0.79–0.91
22C3 VENTANA vs 22C3 DAKO	Pathologist #1	0.74	0.6–0.84	0.85	0.71–0.99	0.83	0.74–0.92
	Pathologist #2	0.70	0.55–0.85	0.96	0.89–1	0.86	0.80–0.93
	Pathologist #3	0.79	0.66–0.92	0.97	0.90–1	0.89	0.82–0.96
	Average scores	0.86	0.75–0.97	0.89	0.77–1.0	0.91	0.86–0.96
SP263 VENTANA vs 22C3 DAKO	Pathologist #1	0.64	0.47–0.8	0.97	0.90–1	0.84	0.77–0.92
	Pathologist #2	0.60	0.43–0.77	0.93	0.84–1	0.82	0.75–0.89
	Pathologist #3	0.79	0.66–0.92	1	NA	0.94	0.92–0.96
	Average scores	0.66	0.5–0.82	0.97	0.9–1.0	0.85	0.78–0.93

Table 3
Light's kappa statistics for agreement between observers.

Assay	1% positivity threshold	50% positivity threshold	1% and 50% thresholds (3 categories)
22C3 VENTANA	0.59	0.88	0.78
22C3 DAKO	0.62	0.96	0.94
SP263 VENTANA	0.86	0.98	0.84

VENTANA assays produce highly concordant PD-L1 scores, in keeping with potentially interchangeable use. The Ventana platform is more widely available than the Dako autostainer [14,15,17], such that demonstration of the analytical concordance of the 22C3 VENTANA assay with commercial assays may have particular utility in expanding the number of sites equipped to performed PD-L1 testing.

The overall percent agreement between 22C3 DAKO and 22C3 VENTANA assays in our study is similar to that in prior studies (83–97.6% for 1% threshold; 87.8–96.7% for 50% threshold) [14,19], as is the κ statistic for 3-category scoring (0.77) [14]. A tendency for slightly lower scores in 22C3 VENTANA versus 22C3 DAKO assays was also noted in two prior studies [14,15]. The small difference in continuous scores in our study, while statistically significant, is likely not clinically significant as it resulted in only 3.5% fewer cases passing the 50% threshold in 22C3 VENTANA versus 22C3 DAKO assays. There were no statistically significant differences between assays when clinically relevant thresholds were employed. Moreover, the percent agreement between the two commercial assays tended to be lower than the percent agreements with the 22C3 VENTANA assay, consistent with the 22C3 VENTANA assay performing at least as well as the commercial assays.

In the present study, the 22C3 DAKO and SP263 VENTANA commercial assays showed agreement similar to that in the Blueprint study [13] and consistent with the substantial concordance of these assays in most studies [23]. Most prior studies support our observation that interobserver and inter-assay agreement is greater at a 50% threshold than a 1% threshold [13,14,23,25]. Interobserver agreement between the three pathologists in our study was similar to that in a study of 27 pathologists (κ = 0.63 for 1% threshold, κ ~ 0.83 for 50% threshold) [26]. Within our study, interobserver and inter-assay agreement were similar, such that interobserver variability may have as much impact on the reliability of scores as the assay itself. Other studies have estimated that inter-observer variability is even greater than the variability between assays [25,27].

We were unable to develop a satisfactory assay using SP263 antibody on the Dako Omnis platform. Attempts were limited by the commercial availability of only pre-diluted antibody, such that antibody concentration could not be optimized. While protocols for SP263 on the Dako Autostainer have been proposed and were considered to have adequate concordance with the SP263 VENTANA assay (κ = 0.83–0.86) [14], no protocols for use of SP263 on the newer and more automated Dako Omnis instrument have yet, to our knowledge, been published. A high failure rate of LDTs for predictive biomarkers is not rare, as recently published in an Italian multicenter study [20].

We note that evidence of true interchangeability requires clinical outcome data for immunotherapy treated patients [4], which is beyond the scope of our study. The technical comparability of 22C3 VENTANA assays supports the notion that clinical assessment of interchangeability (which has not yet been performed) could be worthwhile. Only eight patients whose samples were used in our study had received immunotherapy, reflecting that resection specimens tend to be obtained from patients with early-stage disease, not currently eligible for immunotherapy.

We consider the use of only large resection specimens in our study to be an asset, as such specimens capture a greater spectrum of the heterogeneity in tumor morphology, staining intensity and immune cell infiltration than smaller biopsies. In contrast, two of the prior studies on 22C3 VENTANA LDTs used tissue microarrays rather than whole sections [15,18]. We also note that all slides in our study were independently scored by the same three pathologists, in contrast to prior studies that had only one scoring pathologist [14,18], had all assays on one case scored consecutively (such that interpretation of one assay may be biased by that of another) [14], or had different pathologists score different assays (such that interobserver variability may confound results) [17]. Compared to other LDT's using 22C3 antibody on a Ventana platform, our method uses a shorter cell conditioning step (46 min vs 64 min [15,16,18]), slightly more concentrated 22C3 antibody (1:40 vs 1:50 [14,16]), longer incubation with 22C3 antibody (64 min rather than 60 min [14], 40 min [15], 32 min [16], or 16 min [18]), and cooler temperature for 22C3 antibody incubation (room temperature rather than 37°C [16] or 36 °C [14,18]). Among PD-L1 assays that have shown technical concordance, the assay most suitable for implementation will depend on variables specific to each laboratory. Cost effectiveness of an assay depends on factors including the overall testing volume of the laboratory, specific contracts with various vendors related to instrument and reagent purchasing, and technologists' salaries determined by each institution, all of which are outside the scope of this analysis. Cost may also be influenced by the need for

preparatory steps requiring additional technician time and availability, such as the preparatory steps required for the Dako Autostainer but not the Dako Omnis or Ventana platforms. Waste disposal costs may also differ between platforms; for instance, the Ventana platform mixes reagents that do and not require special disposal, such that the volume requiring special waste handling is greater than for the Dako Omnis and Autostainer platforms.

Overall, we find the 22C3 DAKO, SP263 VENTANA and 22C3 VENTANA assays to have substantial technical concordance, in keeping with potential interchangeable use of these assays. Validation of LDTs for PD-L1, as performed here for our 22C3 VENTANA assay, is a key step in enhancing the accessibility of PD-L1 testing.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.anndiagpath.2020.151590.

Funding

This study received funding from AstraZeneca Canada.

Declaration of competing interest

The authors have no conflicts of interest to declare.

References

[1] Gong J, Chehrazi-Raffle A, Reddi S, Salgia R. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. J Immunother Cancer 2018;6(1):8. https://doi.org/10.1186/s40425-018-0316-z.

[2] Büttner R, Gosney JR, Skov BG, 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(34):3867–76. https://doi.org/10.1200/JCO.2017.74.7642.

[3] Ionescu DN, Downes MR, Christofides A, Tsao MS. Harmonization of PD-L1 testing in oncology: a Canadian pathology perspective. Curr Oncol 2018;25(3):e209–16. https://doi.org/10.3747/co.25.4031.

[4] Cheung CC, Barnes P, Bigras G, et al. Fit-for-purpose PD-L1 biomarker testing for patient selection in immuno-oncology: guidelines for clinical laboratories from the Canadian Association of Pathologists-Association Canadienne Des Pathologistes (CAP-ACP). Appl Immunohistochem Mol Morphol 2019;27(10):699–714. https://doi.org/10.1097/PAI.0000000000000800.

[5] Pai-Scherf L, Blumenthal GM, Li H, et al. FDA approval summary: pembrolizumab for treatment of metastatic non-small cell lung cancer: first-line therapy and beyond. Oncologist 2017;22(11):1392–9. https://doi.org/10.1634/theoncologist.2017-0078.

[6] Herbst RS, Baas P, Kim D-W, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387(10027):1540–50. https://doi.org/10.1016/S0140-6736(15)01281-7.

[7] Reck M, Rodríguez-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):1823–33. https://doi.org/10.1056/NEJMoa1606774.

[8] US Food and Drug Adminstration. FDA expands pembrolizumab indication for first-line treatment of NSCLC (TPS ≥1%) Published online https://www-fda-gov.ezproxy.library.ubc.ca/drugs/fda-expands-pembrolizumab-indication-first-line-treatment-nsclc-tps-1; 2019.

[9] US Food and Drug Adminstration. Durvalumab (Imfinzi) Published online https://www-fda-gov.ezproxy.library.ubc.ca/drugs/resources-information-approved-drugs/durvalumab-imfinzi; 2017.

[10] Botticella A, Mezquita L, Le Pechoux C, Planchard D. Durvalumab for stage III non-small-cell lung cancer patients: clinical evidence and real-world experience. Ther Adv Respir Dis 2019;13. https://doi.org/10.1177/1753466619885530. 1753466619885530.

[11] US Food and Drug Adminstration. FDA expands approval of Imfinzi to reduce the risk of non-small cell lung cancer progressing Published online https://www-fda-gov.ezproxy.library.ubc.ca/news-events/press-announcements/fda-expands-approval-imfinzi-reduce-risk-non-small-cell-lung-cancer-progressing; 2018.

[12] European Medical Agency. Imfinzi: summary of product characteristics Published online https://www.ema.europa.eu/en/documents/product-information/imfizi-epar-product-information_en.pdf; 2018.

[13] 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):1302–11. https://doi.org/10.1016/j.jtho.2018.05.013.

[14] Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann Oncol 2018;29(4):953–8. https://doi.org/10.1093/annonc/mdy014.

[15] Savic S, Berezowska S, Eppenberger-Castori S, et al. PD-L1 testing of non-small cell lung cancer using different antibodies and platforms: a Swiss cross-validation study. Virchows Arch 2019;475(1):67–76. https://doi.org/10.1007/s00428-019-02582-0.

[16] Ilie M, Khambata-Ford S, Copie-Bergman C, et al. Use of the 22C3 anti-PD-L1 antibody to determine PD-L1 expression in multiple automated immunohistochemistry platforms. PLoS One 2017;12(8):e0183023https://doi.org/10.1371/journal.pone.0183023.

[17] Neuman T, London M, Kania-Almog J, et al. A harmonization study for the use of 22C3 PD-L1 immunohistochemical staining on Ventana's platform. J Thorac Oncol 2016;11(11):1863–8. https://doi.org/10.1016/j.jtho.2016.08.146.

[18] Hendry S, Byrne DJ, Wright GM, et al. Comparison of four PD-L1 immunohistochemical assays in lung cancer. J Thorac Oncol 2018;13(3):367–76. https://doi.org/10.1016/j.jtho.2017.11.112.

[19] Munari E, Zamboni G, Marconi M, et al. PD-L1 expression heterogeneity in non-small cell lung cancer: evaluation of small biopsies reliability. Oncotarget 2017;8(52):90123–31. https://doi.org/10.18632/oncotarget.21485.

[20] Vigliar E, Malapelle U, Bono F, et al. The reproducibility of the immunohistochemical PD-L1 testing in non-small-cell lung cancer: a multicentric Italian experience. Biomed Res Int 2019;2019:6832909https://doi.org/10.1155/2019/6832909.

[21] Torlakovic E, Albadine R, Bigras G, et al. Canadian multicentre project on standardization of PD-L1 22C3 immunohistochemistry laboratory developed tests for pembrolizumab therapy in non-small cell lung cancer. J Thorac Oncol 2020. https://doi.org/10.1016/j.jtho.2020.03.029. Published online April 15.

[22] 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(2):208–22. https://doi.org/10.1016/j.jtho.2016.11.2228.

[23] Koomen BM, Badrising SK, van den Heuvel MM, Willems SM. Comparability of PD-L1 immunohistochemistry assays for non-small-cell lung cancer: a systematic review. Histopathology 2019. https://doi.org/10.1111/his.14040. Published online December 2.

[24] Widmaier M, Wiestler T, Walker J, et al. Comparison of continuous measures across diagnostic PD-L1 assays in non-small cell lung cancer using automated image analysis. Mod Pathol 2019. https://doi.org/10.1038/s41379-019-0349-y. Published online September 16.

[25] 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(10):1411–21. https://doi.org/10.1038/modpathol.2017.59.

[26] Chang S, Park HK, Choi Y-L, Jang SJ. Interobserver reproducibility of PD-L1 biomarker in non-small cell lung cancer: a multi-institutional study by 27 pathologists. J Pathol Transl Med 2019;53(6):347–53. https://doi.org/10.4132/jptm.2019.09.29.

[27] 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(14):3585–91. https://doi.org/10.1158/1078-0432.CCR-16-2375.