Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non–Small Cell Lung Cancer

Marianne J. Ratcliffe1, Alan Sharpe2, Anita Midha1, Craig Barker2, Marietta Scott2, Paul Scorer2, Hytham Al-Masri3, Marlon C. Rebelatto4, and Jill Walker2

Abstract

Purpose: Immunotherapies targeting programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1) demonstrate encouraging antitumor activity and manageable tolerability in non–small cell lung cancer (NSCLC), especially in patients with high tumor PD-L1 expression, as detected by companion or complementary diagnostic assays developed for individual agents. A laboratory is unlikely to use multiple assay platforms. Furthermore, commercially available diagnostic assays are not standardized, and different assay methods could lead to inappropriate treatment selection. This study establishes the extent of concordance between three validated, commercially available PD-L1 IHC diagnostic assays for NSCLC patients [Ventana SP263 (durvalumab), Dako 22C3 (pembrolizumab), and Dako 28-8 (nivolumab)].

Experimental Design: Five hundred formalin-fixed, paraffin-embedded archival NSCLC samples were obtained from commercial sources. Stained slides were read in batches on an assay-by-assay basis by a single pathologist trained in all methods, in a Clinical Laboratory Improvements Amendments program–certified laboratory. An additional pathologist performed an independent review of 200 stained samples for each assay.

Results: PD-L1 expression was evaluable with all assays in 493 samples. The three assays showed similar patterns of tumor membrane staining, with high correlation between percent PD-L1 staining. An overall percentage agreement of >90% was achieved between assays at multiple expression cutoffs, including 1%, 10%, 25%, and 50% tumor membrane staining.

Conclusions: This study builds optimism that harmonization between assays may be possible, and that the three assays studied could potentially be used interchangeably to identify patients most likely to respond to anti-PD-1/PD-L1 immunotherapies, provided the appropriate clinically defined algorithm and agent are always linked. Clin Cancer Res; 1–7. ©2017 AACR.

Introduction

Some tumors can evade detection by the immune system by exploiting inhibitory checkpoint pathways that suppress antitumor T-cell responses (1). Among the most important of these checkpoints is the programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1) pathway, in which PD-L1 expressed by tumor or tumor-infiltrating immune cells binds to PD-1, inhibiting T-cell receptor signaling and blocking the antitumor immune response (2-4). Blocking antibodies that target PD-1 or PD-L1 have been developed to interrupt this interaction (2), and a number of effective therapeutics are emerging in multiple tumor types, including non–small cell lung cancer (NSCLC; refs. 3, 4). For example, the anti-PD-1 antibodies pembrolizumab and nivolumab are clinically active in patients with NSCLC and achieve improved responses in patients with high tumor PD-L1 expression compared with those expressing low or no tumor PD-L1 (5-8). Pembrolizumab is approved for use in patients with metastatic NSCLC whose tumors express PD-L1 in the membrane of ≥50% of tumor cells, as determined by an FDA-approved test, and who have disease progression on or after prior therapy (9). The companion diagnostic approved for use with pembrolizumab in NSCLC is Dako PD-L1 IHC 22C3 pharmDx (10). Nivolumab is approved for use in patients with metastatic NSCLC whose tumors have progressed on or after prior therapy (11, 12). PD-L1 testing is not required for the use of nivolumab in NSCLC; however, nivolumab does have an FDA-approved complementary diagnostic (Dako PD-L1 IHC 28-8 pharmDx) (13).

Atezolizumab is a PD-L1 inhibitor that has recently received FDA approval for use in patients with metastatic NSCLC whose disease progressed during or following platinum-containing chemotherapy (14) along with the VENTANA PD-L1 (SP142) Assay as a complementary diagnostic (15). At the time of initiation of this study, the VENTANA PD-L1 (SP142) Assay was not
Responses to anti-PD-1 and anti-PD-L1 agents are enhanced in patients whose tumors express high levels of PD-L1 compared with those expressing low or no tumor PD-L1. As such, reliable testing assays are required to inform treatment choices. Broad access to high-quality PD-L1 testing will help clinicians to identify the most appropriate treatment option for individual patients with NSCLC, including chemotherapy versus anti-PD-1/PD-L1, and single-agent anti-PD-1/PD-L1 versus combination immunotherapy regimens. Our study demonstrates analytic equivalence of tumor membrane staining between three different validated, commercially available PD-L1 IHC diagnostic assays. These results build confidence that any testing laboratory with access to one of these high-quality tests will be able to provide information relevant to clinical decision making with anti-PD-1 and anti-PD-L1 therapeutics, provided the correct assay cutoff is used.

Translational Relevance

Responses to anti-PD-1 and anti-PD-L1 agents are enhanced in patients whose tumors express high levels of PD-L1 compared with those expressing low or no tumor PD-L1. As such, reliable testing assays are required to inform treatment choices. Broad access to high-quality PD-L1 testing will help clinicians to identify the most appropriate treatment option for individual patients with NSCLC, including chemotherapy versus anti-PD-1/PD-L1, and single-agent anti-PD-1/PD-L1 versus combination immunotherapy regimens. Our study demonstrates analytic equivalence of tumor membrane staining between three different validated, commercially available PD-L1 IHC diagnostic assays. These results build confidence that any testing laboratory with access to one of these high-quality tests will be able to provide information relevant to clinical decision making with anti-PD-1 and anti-PD-L1 therapeutics, provided the correct assay cutoff is used.

The pembrolizumab prescribing information states that PD-L1 expression should be determined using an FDA-approved test (9): Dako PD-L1 IHC 22C3 pharmDx was used in the pivotal KEYNOTE 001 study (7). Despite the link between drug and diagnostic approval, the need to drive efficiency in patient testing by preserving tissue and resources makes it likely that some patients might be screened for pembrolizumab using other assays such as Dako PD-L1 IHC 28-8 pharmDx or an LDT. This raises potential concerns, as pembrolizumab approval was based on clinical outcome data derived using Dako 22C3 with a cutoff of PD-L1 expression in \( \geq 50\% \) of tumor cells, whereas Dako 28-8 has been validated using cutoffs of \( \geq 1\% \), \( \geq 5\% \), and \( \geq 10\% \) of tumor cells, each predicting clinical benefit with nivolumab. Dako 28-8 has not been validated at the \( \geq 50\% \) cutoff, the value associated with clinical benefit of pembrolizumab. There is potential harm to patients if tests and drugs are cross matched by end users without clear evidence of concordance (27). An added complication arises from the fact that PD-L1 status in NSCLC determined by the VENTANA PD-L1 (SP142) Assay is based on either the percentage of PD-L1-expressing tumor cells or the proportion of tumor area occupied by PD-L1-expressing tumor-infiltrating immune cells (22).

It is imperative to be able to compare the analytic performance of PD-L1 diagnostic assays to allow appropriate interpretation of their use with respect to treatment selection. Comparisons must be performed according to validated protocols, as directed by the manufacturer and with appropriate training. The current study was undertaken to establish the extent of analytic concordance between the VENTANA PD-L1 (SP263) Assay and the two validated PD-L1 IHC diagnostic assays that have been granted FDA approval in NSCLC at the time of the current analysis (Dako 22C3 and Dako 28-8), as a first step toward assay harmonization.

Materials and Methods

Tumor samples

For this analysis, 500 archival NSCLC samples (formalin-fixed, paraffin-embedded blocks) were obtained from commercial sources (Asterand; ProteoGenex; Tissue Solutions). Sample age ranged from 1 to 4 years, based on the date of excision. Fresh sections were cut for the purpose of the study. The final dataset comprised 493 samples; 7 samples were excluded from the final analysis as they did not meet sample requirements. Consecutive sections were used to reduce variability between assays due to tumor heterogeneity. All samples were read in batches on an assay-by-assay basis, by a single pathologist trained by the manufacturers, in a Clinical Laboratory Improvement Amendments (CLIA) program–certified laboratory (Hematogenix). Samples were assessed and scored as per package inserts provided by Ventana and Dako; assessment was therefore based on tumor cell membrane staining, not immune cell staining. Details of the scoring for each assay are provided in Supplementary Table S1. A washout period was included between reads of samples from the same patient to remove bias. The average washout periods were: 29 days for Ventana SP263/Dako 22C3, 3 days for Ventana SP263/Dako 28-8, and 26 days for Dako 28-8/Dako 22C3. Subsequently, the samples stained with SP263 were additionally scored at the 1%, 5%, and 10% PD-L1 expression cutoff levels (not as per package insert).
to allow direct comparison of these cutoffs with the other two assays.

An independent review of 200 samples, representing the dynamic range of PD-L1 expression, was performed by an additional pathologist for each of the assays (Ventana SP263, Dako 22C3, and Dako 28-8) to provide an assessment of the composite variability observed between different assays and observers. Washout periods were included between each case and each assay to prevent bias. Interobserver analysis between the original and second reader was performed.

Analysis plan
Overall percentage agreement (OPA) was calculated pairwise between assays at multiple PD-L1 expression cutoff values. For prespecified and clinically relevant expression cutoffs, negative percentage agreement (NPA) and positive percentage agreement (PPA) were calculated (28). For each metric, a lower boundary of the 95% confidence interval (CI) was calculated with no upper bound using the Clopper–Pearson method (29). OPA between the study pathologist and the independent pathologist was calculated at the clinically relevant 1%, 10%, 25%, and 50% PD-L1 expression cutoff levels. An OPA >90% is typical within-assay agreement for IHC (30).

Results
Sample demographics
The demographics of the 493 patients who provided NSCLC samples included in the analysis are shown in Supplementary Table S3. Samples were from patients with stage I–IV NSCLC; the majority (75.3%) of patients were Caucasian and approximately half each had squamous and nonsquamous histology.

PD-L1 membrane staining
All three PD-L1 assays showed similar patterns of staining (Fig. 1). Correlations between assays are shown in Fig. 2; the Spearman correlation coefficients were all >0.9, indicating high associations between the three assays.

PD-L1 expression prevalence
PD-L1 expression on tumor cell membrane displayed a continuous range from 0% to 100%. Binned data are presented in Fig. 3.

OPA
An OPA of >90% was achieved between assays at multiple PD-L1 expression cutoffs (Table 1; Supplementary Table S4), which is typical within-assay agreement for IHC (30).

PPA and NPA
PPA and NPA were determined relative to different reference assays at their clinically relevant PD-L1 expression cutoffs. Hence, Dako 28-8 at the 1% and 10% cutoffs, Ventana SP263 at the 25% cutoff, and Dako 22C3 at the 50% cutoff were set as the reference assays for the comparisons against the other assays at matching cutoffs (Table 2). A high degree of agreement was observed between assays at the different cutoffs, with NPA and PPA >85% for each comparison.

Independent pathology review
In a subset of 200 samples, good agreement between the original CLIA laboratory study pathologist results and independent pathologist review was observed for all three assays. OPA at cutoffs of 10%, 25%, and 50% PD-L1 expression was >85%. OPA was slightly lower at the 1% cutoff (range 75.9%–77.0%; Table 3).

Discussion
It is now understood that PD-L1 protein expression is not binary; it is a continuous variable whereby a range of PD-L1 expression levels are observed, with tumor heterogeneity also frequently noted (21, 23, 31), all of which may have biological and/or clinical significance. Staining cutoffs are used to classify patients with high or low/no tumor PD-L1 expression (20, 23), and the cutoff appropriate for clinical decision making for each PD-1 or PD-L1 therapy is determined by clinical data.

Expression cutoffs for PD-L1 used in clinical trials to date vary between 1% and 50%, depending on the study and assay used (refs. 5, 7, 17, 19; Supplementary Table S1). Two recent reviews have examined the correlation between PD-L1 expression and outcome in patients with NSCLC treated with anti-PD-1 or anti-PD-L1 therapeutic agents (32, 33). One of these, a meta-analysis by Abdel-Rahman, found that the advantage of anti-PD-1 agents over docetaxel in second-line treatment of NSCLC is limited to tumors with >1% PD-L1 expression, and that the benefit of anti-PD-1 or anti-PD-L1–targeted agents in terms of objective response rate may have a positive correlation with the intensity of PD-L1 staining (32). The recent KEYNOTE-010 study compared pembrolizumab with docetaxel in previously treated patients with NSCLC, using 1% and 50% cutoffs for PD-L1 expression (34). Significant superiority over docetaxel in overall survival and objective response rate was seen at both expression cutoffs, with a trend for higher median overall survival in patients with ≥50% tumor PD-L1 expression.
Ventana SP263 versus Dako 22C3
Spearman correlation coefficient = 0.925

Ventana SP263 versus Dako 28-8
Spearman correlation coefficient = 0.948

Dako 28-8 versus Dako 22C3
Spearman correlation coefficient = 0.954

Figure 2.
Correlation of percentage tumor cell membrane staining in non–small cell lung cancer samples.

Table 1. OPA between assays at multiple expression cutoff levels

<table>
<thead>
<tr>
<th>Expression cutoff</th>
<th>Ventana SP263 vs. Dako 28-8 OPA (lower 95% CI), %</th>
<th>Dako 22C3 vs. Dako 28-8 OPA (lower 95% CI), %</th>
<th>Ventana SP263 vs. Dako 22C3 OPA (lower 95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1%</td>
<td>91.7 (89.3)</td>
<td>93.7 (91.7)</td>
<td>91.1 (88.7)</td>
</tr>
<tr>
<td>≥10%</td>
<td>92.9 (90.7)</td>
<td>94.9 (93.0)</td>
<td>92.7 (90.5)</td>
</tr>
<tr>
<td>≥25%</td>
<td>94.9 (93.0)</td>
<td>96.6 (94.9)</td>
<td>94.3 (92.3)</td>
</tr>
<tr>
<td>≥50%</td>
<td>95.9 (94.2)</td>
<td>97.2 (95.6)</td>
<td>93.5 (91.4)</td>
</tr>
</tbody>
</table>
versus patients with ≥1% expression. Increased progression-free survival versus docetaxel was observed at both expression cutoffs, but the difference was only statistically significant in patients with ≥50% tumor PD-L1 expression. On the basis of these results, the FDA expanded the indication for pembrolizumab in the second-line treatment of NSCLC to include all patients with tumor PD-L1 expression (≥1% expression cutoff; ref. 35). Thus, the choice of cutoff needs to balance the drive for higher efficacy in patients selected for treatment versus the opportunity to benefit the highest number of patients. The appropriate PD-L1 expression cutoff may also depend on the particular therapy used and needs to be determined by robust clinical data. Multiple PD-L1 expression cutoffs are therefore likely to be applied to different therapy selections.

It is anticipated that pathologists will need to drive efficiencies in testing and that substituting one staining assay for another will become widespread. There is a potential risk to patients if the perceived issues with assay variability are borne out. Being able to use different assays that give similar outcomes would be of tremendous value for pathology laboratories, requiring each laboratory to employ only one of the testing platforms.

The International Association for the Study of Lung Cancer and the American Association for Cancer Research (AACR) have joined together with pharmaceutical companies and diagnostics providers to test the technical equivalency (though not the predictive equivalency) of four PD-L1 IHC assays (Dako 28-8, Dako 22C3, Ventana SP263, and Ventana SP142; refs. 27, 36). The initial phase I part of this investigation was a feasibility study on a small cohort of 38 noninterventional NSCLC cases, each stained using all four assays and scored by three pathologists (36). Regarding tumor cell staining, Dako 22C3, Dako 28-8, and Ventana SP263 demonstrated similar analytic performance, whereas Ventana SP142 consistently labeled fewer tumor cells (36). The International Association for the Study of Lung Cancer will conduct a larger phase II study looking at different sample types, and the effects of mixing antibodies and platforms.

Another recent study assessed interobserver concordance and PD-L1 IHC staining patterns in 15 pulmonary carcinoma resection specimens using three assays: Dako 28-8, Ventana SP142, and Ventana SP263 (37). Data showed that carcinoma cells could be reproducibly scored, with no differences in interobserver concordance among the tested assays. The scoring of immune cells yielded low concordance rates and the authors concluded that immune cell scoring might require specific standardization.

We did not assess the VENTANA PD-L1 (SP142) Assay in the current study because the validated assay was not commercially available at the time of analysis. As discussed previously, the VENTANA PD-L1 (SP142) Assay recently received FDA approval as a complementary diagnostic for atezolizumab in patients with previously treated metastatic NSCLC (15).
The current study (the largest to date, with 493 tumor biopsy samples from patients with NSCLC) demonstrated a high analytic correlation between the three commercially available PD-L1 assays used by AstraZeneca, Bristol-Myers Squibb, and Merck Sharp & Dohme in determining PD-L1 protein expression, with high levels of agreement between assays at multiple matched expression cutoff points. Preliminary results from this analysis were presented at AACR (38). Subsequently, a further definition of scoring of the samples with the VENTANA PD-L1 (SP263) Assay to include <1%, 1%–4%, 5%–9%, and ≥10% cutoffs (currently outside the "trained" scoring range for this assay) was performed to allow a more direct comparison with the two Dako assays. The assays have closely aligned dynamic ranges. An OPA of >90% (range 91.1%–97.2%), which is typical within-assay agreement for IHC (30), was achieved across the dynamic range at multiple PD-L1 expression cutoffs. PPA among the three assays (range 86.0%–97.5%) varied more than NPA (range 93.5%–98.8%) and OPA, according to the expression cutoffs used for the comparator assays. It is possible to optimize for either NPA or PPA by selecting different cutoffs for the comparison, which would facilitate the comparison of results from clinical studies that have used different tests, but such optimization may come at the expense of OPA. For Dako PD-L1 IHC 28-8 pharmDx, the 10% expression cutoff was selected in addition to the 1% cutoff for evaluation as a reference value because the CheckMate 057 study showed that second-line nivolumab offered no significant efficacy advantage over docetaxel in patients with NSCLC harboring tumors with <10% PD-L1 staining, as measured by Dako 28-8 (5).

The independent pathologist’s review of 200 samples was largely consistent with that of the CLIA laboratory study pathologist, supporting the validity of the study results. Interestingly, the variability between different pathologists scoring the same stained samples appeared higher than the variability between different assays scored by a single reader. Concordance between different pathologists scoring the same slides was lower for samples with staining below 10%, indicating that reader interpretation may be more important in assays that use lower cutoffs. Together, these data indicate that the intrinsic assay variability is lower than that between readers, supporting our conclusions that the assays themselves can be considered effectively equivalent in terms of tumor cell membrane staining. Effective training of pathologists at specific cutoffs, particularly if the cutoff is low, will be needed to ensure reproducibility of patient classification across different laboratories.

Our study builds optimism that harmonization between assays is possible, and that the determination of PD-L1 staining levels by any of the three assays studied could be used interchangeably as an aid to therapeutic decision making, provided that the appropriate, clinically validated cutoffs for each of the respective anti-PD-1 or anti-PD-L1 therapeutic antibodies are applied. Despite the fact that the three antibodies are different clones developed for use with different assays, they showed strong agreement; however, it should be noted that clones other than those examined herein might not agree as closely. In addition, these results should not be extrapolated to antibody clones embedded in other assay protocols unless these protocols are fully validated, and should also not be extrapolated to other cancer indications without a separate study. A recent, smaller scale study assessing the performance of different PD-L1 antibodies indicated good concordance between the 22C3, 28-8, and SP263 antibodies in detection of PD-L1 in cell lines and tumor samples (39), consistent with our study. In that study (39), locked assays were not used, and concordance depended on the staining protocol used, emphasizing the importance of ensuring all elements of an IHC protocol are properly validated.

PD-L1 expression is a continuous variable, and a simplistic definition of PD-L1 status as positive or negative will not provide the flexibility needed to inform treatment decisions in an environment with an increasing number of available anti-PD-1/PD-L1 therapies. Findings from this study of three commercially available assays indicate that it may be possible to extrapolate the PD-L1 staining level results from one assay to another, driving efficiency by allowing the tests to be used interchangeably to identify patients with the greatest likelihood of responding to anti-PD-1 or anti-PD-L1 immunotherapeutic agents, providing the appropriate, clinically defined algorithm or cutoff is applied to each drug. Further discussions within the pathology and clinical community are required to agree the appropriate route to meet this goal in practice.

Disclosure of Potential Conflicts of Interest
M.J. Ratcliffe and M. Scott hold ownership interest (including patents) in AstraZeneca. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: M.J. Ratcliffe, A. Sharpe, A. Midha, C. Barker, P. Scorer, J. Walker
Development of methodology: A. Sharpe, A. Midha, C. Barker, P. Scorer, J. Walker
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Midha, C. Barker, H. Al-Masri
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.J. Ratcliffe, A. Sharpe, A. Midha, C. Barker, M. Scott, H. Al-Masri, J. Walker
Writing, review, and/or revision of the manuscript: M.J. Ratcliffe, A. Sharpe, A. Midha, C. Barker, M. Scott, P. Scorer, H. Al-Masri, M.C. Rebelatto, J. Walker
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Sharpe, C. Barker, H. Al-Masri
Study supervision: M.J. Ratcliffe, A. Midha, C. Barker, P. Scorer, H. Al-Masri

Grant Support
This study was sponsored by AstraZeneca. Medical writing and editorial assistance were provided by Abigail Marmont of CircleScience, an Ashfield company (funded by AstraZeneca).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 23, 2016; revised December 9, 2016; accepted December 30, 2016; published OnlineFirst January 10, 2017.

References
27. PD-L1 Blueprint Project. Available from: http://www.nationalacademies.edu/hmd/}
Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non–Small Cell Lung Cancer

Marianne J. Ratcliffe, Alan Sharpe, Anita Midha, et al.

Clin Cancer Res Published OnlineFirst January 10, 2017.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-16-2375

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2017/01/10/1078-0432.CCR-16-2375.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.