Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non–Small-cell Lung Cancer

Charlotte Roach, BS,* Nancy Zhang, MD,* Ellie Corigliano, PhD,* Malinka Jansson, MS,* Grant Toland, BS,* Gary Ponto, MD,* Marisa Dolled-Filhart, PhD,† Kenneth Emancipator, MD,† Dave Stanforth, MBA,* and Karina Kulangara, PhD*

Abstract: A companion diagnostic assay was codeveloped by Dako for pembrolizumab non–small-cell lung cancer clinical trials to detect PD-L1 expression by immunohistochemistry (IHC). This automated IHC assay has been analytically verified and validated using Dako’s autostainer Link 48 and 22C3 mouse anti-PD-L1 monoclonal antibody to detect the PD-L1 expression in formalin-fixed paraffin-embedded human tumor tissue specimens. The PD-L1 22C3 IHC assay was optimized for high sensitivity and specificity. Repeatability and reproducibility studies were conducted at Dako and at 3 Clinical Laboratory Improvement Amendments certified laboratories during assay development. The studies included: interobserver and intraobserver, interinstrument, interoperator, interday, and intralot, and intraday and intrarun. All precision studies performed at Dako and external laboratories achieved >85% point-estimate agreements for all 3 agreement types (negative, positive, and overall). A clinical cutoff (tumor proportion score ≥50%) of PD-L1 expression was determined and evaluated through a phase 1 clinical trial (KEYNOTE-001) for advanced non–small-cell lung cancer patients treated with pembrolizumab. The treatment effect of pembrolizumab in the 61 subjects who had a tumor PD-L1 of tumor proportion score ≥50% was substantial, with an overall response rate of 41% (95% confidence interval, 28.6-54.3) as compared with 20.6% (95% confidence interval, 15.5-26.5) observed in the 223 subjects irrespective of PD-L1 status. PD-L1 IHC 22C3 pharmDx is a sensitive, precise, and robust companion diagnostic assay, which will facilitate safe and effective use for pembrolizumab in cancer patients.

Key Words: programmed cell death 1, non–small-cell lung cancer, immunohistochemistry

An intact immune system is capable of recognizing and eliminating tumor cells through immune check points. However, increasing evidence exists that tumors can evade adaptive immunity by utilizing T-cell check point pathways.1 Programmed cell death 1 (PD-1) is a negative costimulatory receptor expressed primarily on the surface of activated T cells.2,3 PD-L1 and PD-L2, the PD-1 ligands, can be expressed on the surface of tumor cells. The binding of PD-1 and its ligands is a key pathway exploited by tumors to suppress immune control.4,5 The expression of PD-L1 has been reported in a number of human malignancies. In patients with non–small-cell lung cancer (NSCLC), PD-L1 expression appears to be associated with poor prognosis.6 In clinical trials, anti-PD-1 and anti-PD-L1 antibodies produce durable responses in approximately 20% of unselected patients with NSCLC.7–10 Preliminary molecular marker studies showed the correlation of PD-L1 expression in tumor or inflammatory cells, in pretreatment tumor biopsies, with clinical outcomes, which indicate that PD-L1 may be a predictive biomarker in a subgroup.2,7,11,12 However, developing a reliable and validated biomarker assay that identifies patients with an increased probability of response to anti-PD-1 or anti-PD-L1 therapies remains a challenge.

Pembrolizumab, a highly selective, humanized monoclonal immunoglobulin G4 kappa isotype antibody against PD-1 developed by Merck & Co. has demonstrated antitumor efficacy in phase I clinical trials (KEYNOTE-001) for patients with advanced NSCLC and highly expressing PD-L1 tumors.13 This novel targeted therapy necessitates the availability of a high-quality diagnostic biomarker to facilitate its safe and effective use.14,15

Immunohistochemistry (IHC) assays using different primary antibodies, and antibody-specific scoring approaches have been reported to assess the prevalence of PD-L1 positivity in NSCLC. The PD-L1 positivity rate by scoring the staining of PD-L1 on membrane and cytoplasm...
from different NSCLC cohorts varied from 20% to 50%. Multiple factors may contribute to the varied PD-L1 prevalence, including differences in antibodies, assay methods, stages of the tumors, and treatments before sample collection. The Dako PD-L1 IHC 22C3 pharmDx, an IHC assay using monoclonal antibody 22C3, has been fully developed and used to determine PD-L1 expression in a clinical phase 1 trial (KEYNOTE-001) for patients with advanced NSCLC. The trial has shown that ≥ 50% of PD-L1 expression in tumor cells correlates with significantly improved efficacy of pembrolizumab.

Dako PD-L1 IHC 22C3 pharmDx assay is the first companion diagnostic (cdx) assay for PD-L1 with approval in the United States. In this paper, we present the analytical and clinical validation for Dako PD-L1 IHC 22C3 pharmDx assay, demonstrating its high sensitivity, repeatability, and reproducibility. In addition, the clinical validation in KEYNOTE-001 confirmed this assay as an aid in identifying patients with NSCLC who are eligible for the treatment with pembrolizumab using 50% tumor proportion score (TPS) as a cutoff.

MATERIALS AND METHODS

PD-L1 IHC 22C3 pharmDx Assay

IHC staining procedure was performed using the Dako Autostainer Link 48 platform and an automated staining protocol validated for the PD-L1 IHC 22C3 pharmDx assay. Deparaffinization, rehydration, and target retrieval was performed in the PT Link (Dako PT100) using a 3-in-1 procedure. After incubation with the monoclonal mouse anti-human PD-L1 antibody, clone 22C3 or the negative control reagent, mouse immunoglobulin G isotype control, specimens were incubated with anti-mouse linker antibody specific to the host species of the primary antibody, and then were incubated with a ready-to-use visualization reagent consisting of secondary antibody molecules and horseradish peroxidase molecules coupled to a dextran polymer backbone. The enzymatic conversion of the subsequently added 3,3'-dianinobenzidine tetrahydrochloride chromogen followed by 3,3'-dianinobenzidine tetrahydrochloride enhancer resulted in precipitation of a visible reaction product at the site of antigen. The specimens were then counterstained with hematoxylin and cover-slipped. Results were interpreted using a light microscope. For the development of the finalized assay the sensitivity was optimized with minimum nonspecific staining by adjusting primary antibody concentration and reagent incubation times.

Tissue Specimen Preparation

All specimens used in these studies were formalin-fixed paraffin-embedded (FFPE). Sections were cut at 4 μm thickness, placed on positively charged slides, and dried in an oven for 1 hour. Human tissue sections were dried at 56 to 60°C and experimental cell lines were dried at 40 to 44°C. The cut sections were stored in dark at 2 to 8°C and stained with the PD-L1 IHC assay within 6 months.

Scoring Interpretation

For determination of PD-L1 protein expression, positivity was defined as complete circumferential or partial cell membrane staining of viable tumor cells with 1+ to 3+ intensity. Nonspecific staining was recorded on a 0 to 3 intensity scale, in 0.25 grade increments. Tumor-associated immune cells were excluded from PD-L1 scoring. Cytoplasmic staining, if present, was excluded from the scoring. Scoring was recorded as percentage of PD-L1-positive tumor cells over total tumor cells in the denominator (TPS). NSCLC specimens stained with the negative control reagent must have 0 specific membrane staining and ≤ 1+ intensity nonspecific (nonmembrane) staining.

Phase 1 Clinical Trial

Patients with advanced NSCLC were enrolled in a multicenter, open-label, phase 1 trial (KEYNOTE-001) for treatment with pembrolizumab on the basis of the PD-L1 expression. The major efficacy outcome measures were overall response rate (according to RECIST 1.1 as assessed by blinded-independent central review) and duration of response. One hundred forty-six NSCLC patient samples were retrospectively tested with the PD-L1 IHC 22C3 pharmDx assay to develop the scoring method and to lock down the cutoff. Clinical outcome and PD-L1 expression from another set of 61 patients were then used to validate the predetermined cutoff.

Statistical Analysis

Agreement Calculations for Reproducibility Studies

Average percent negative agreement (ANA), average percent positive agreement (APA), and overall percent agreement (OA) were calculated for intersite and intrasite, and interobserver and intraobserver reproducibility, using 2-sided 95% percentile bootstrap confidence intervals (CI). The CIs were calculated by resampling on specimen IHC status; bootstrap samples were produced by separately resampling positive and negative specimens with replacement. Acceptance criteria were defined as: ANA, APA, and OA of at least 85% at the lower bound of a 2-sided 95% percentile bootstrap CI.

Agreement Calculations for Precision (Repeatability) Studies

All precision studies resulted in 100% agreement. Therefore, the Wilson Score method was used to calculate CIs for the negative agreement (NPA), positive agreement (PPA), and OA. Specifically, CIs were calculated based on the number of independent pair-wise comparisons in each study. Note that the bootstrap method is only applicable for results with some discordance.
RESULTS

Assay Sensitivity

The prevalence of PD-L1 protein in FFPE tissue sections using Dako’s PD-L1 IHC 22C3 pharmDx assay was evaluated. One hundred twenty-seven unique archived NSCLC specimens were stained for PD-L1 with the clone 22C3 monoclonal antibody. All specimens were represented as whole tissue sections, and primary, metastatic, stage III, and stage IV disease specimens were included. PD-L1 expression was observed on tumor and immune cells. For evaluation, PD-L1 expression was only considered to be positive on tumor cell membranes. Partial tumor cell membrane staining was sufficient and a minimum of 100 viable tumor cells needed to be present for an evaluable interpretation. PD-L1 expression was quantified as TPS:

$$\text{TPS} = \left( \frac{\text{Thenumber of viable tumor cells positive for PD-L1}}{\text{total number of viable neoplastic cells}} \right) \times 100\% .$$

PD-L1 expression in tumor tissue bank samples was observed over the entire dynamic staining intensity range (Figs. 1A–C). Seventy-three of 127 (57.5%) specimens did not express PD-L1 (TPS of 0) (Figs. 1A, B); 30 of 127 (23.6%) expressed PD-L1 between 1% and 49% (Fig. 1B), and 23 of 127 (18.4%) expressed PD-L1 with a TPS of >50% (Fig. 1B). This prevalence is similar to the one observed in the clinical protocol 001F.13

Clinical Cutoff Validation

PD-L1 IHC 22C3 pharmDx assay was retrospectively used to investigate the correlation of PD-L1 status with the clinical outcome in a phase 1 clinical trial in which the safety and efficacy of pembrolizumab in 223 patients with advanced NSCLC were assessed. A total of 146 subjects received at least 1 dose of pembrolizumab, had baseline measurable disease by investigator assessed immune-related response criteria and had samples evaluable by PD-L1 IHC 22C3 pharmDx assay. A cutoff was determined using receiver-operating characteristic analysis based on investigator immune-related response criteria in these 146 subjects with ≥19 weeks follow-up. Details of this analysis have been described elsewhere.16 The optimal cutoff was TPS ≥50% for the PD-L1 IHC 22C3 pharmDx assay.17 After the cutoff was determined, clinical outcome data and PD-L1 expression results from another set of 61 previously treated patients were used to analyze the correlation of clinical outcome with PD-L1 status with TPS ≥50%. The treatment effect of pembrolizumab in the 61 previously treated subjects who had a tumor PD-L1 of TPS ≥50% was substantial, with an overall response rate of 41% (95% CI, 28.6-54.3) as compared with 20.6% (95% CI, 15.5-26.5) observed in the 223 subjects irrespective of PD-L1 status (Table 1). The duration of the response, based on patients with a confirmed response (n = 25) by an independent review, was 2.1+, 9.2+ median months. The endpoint has not yet been reached. In 84% of the patients, the response is still ongoing, including 11 patients with ongoing responses of 6 months or longer.

Assay Repeatability and Reproducibility

High levels of repeatability and reproducibility in normal day to day testing across laboratories ensure accuracy of any companion diagnostics. PD-L1 IHC 22C3 pharmDx was tested for repeatability and reproducibility within Dako and across 3 external Clinical Laboratory Improvement Amendments. Sixteen NSCLC tissues were selected for Dako’s internal analytical precision studies. Four tissues (25%) were chosen to be near the TPS 50% diagnostic cutoff in the PD-L1 expression range of 40% to 60% TPS. As part of the analytical validation conducted at Dako, 6 instruments, 6 operators, 6 nonconsecutive days, 3 lots of reagents, and 6 replicates (intrarun repeatability) were used for the studies. The stained slides were assessed for their concordance in positivity and negativity of PD-L1 expression using the clinically relevant cutoff of 50% TPS, with a TPS of ≥50% being considered diagnostically positive, and a TPS <50% being considered diagnostically negative. Statistical analysis was done using the Wilson Score method. For interinstrument, interoperator, interday, and interlot, and intraday and intrarun, 100% agreement was achieved for all 3 agreement endpoints (NPA, PPA, and OA), demonstrating high repeatability (Table 2).

Intersite reproducibility was conducted at 3 external testing laboratories located in the United States and Europe. Each laboratory performed automated staining runs on each of 5 nonconsecutive days. The same set of specimens was stained 5 times, with each set consisting of single tissue sections from 36 FFPE human NSCLC specimens. Efforts were made to select a balanced number of positive and negative specimens with approximately 25% of specimens near the cutoff. The staining was performed by 1 operator (laboratory technician) over a period of at least 20 days, using reagents from 1 PD-L1 22C3 pharmDx lot, on 1 or 2 Autostainer Link 48 instruments at each site. Staining results were evaluated using TPS with a 50% cutoff (as previously described) by 1 observer (pathologist) at each of the 3 testing laboratories. An overview of the study design is provided in Figure 2A. Statistical analysis was done using a percentile bootstrap method. Pair-wise comparisons resulted in 2700 total concordant outcomes (1477 PD-L1 negative and 907 PD-L1 positive), with a total of 316 discordant outcomes. The following agreements were attained: ANA achieved 90.3% with a lower bound of 84.4%, APA achieved 85.2% with a lower bound of 75.6%, and OA achieved 88.3% with a lower bound of 81.4% (Table 3).

Intrasite reproducibility was also evaluated by testing concordance within each site, across the 5 days. Pair-wise comparisons resulted in 1080 total concordant outcomes (601 negative and 373 positive), with a total of 106 discordant outcomes. Analysis results yielded ANA of 91.9% with a lower bound of 88.8%, APA of 87.6%
with a lower bound of 82.5%, and OA of 90.2% with a lower bound of 86.3%.

Testing for reproducibility of interpretation was conducted through a blinded, randomized study at the 3 external testing laboratories. One observer (pathologist) at each of the 3 testing laboratories performed 3 independent interpretations of slides that were previously stained with PD-L1 IHC 22C3 pharmDx. The sample set comprised a total of 62 FFPE human NSCLC specimens that represented the full range of PD-L1 expression, and included PD-L1-positive and PD-L1-negative specimens with respect to the diagnostic cutoff of 50% TPS. Efforts were made to select an equal distribution of positive and negative specimens and approximately 25% of the specimens near the cutoff. An overview of the study design is given in Figure 2B for interobserver variation. To minimize bias in the evaluation process, the slides were evaluated in a specified, randomized order with unique, additional wildcard slides added to each of the 3 reads. A washout period of at least 3 days between the 3 evaluations was performed to minimize observer memory of previous assessment.

Similar to the intersite study, statistical analysis for the observer study was done using a percentile bootstrap method. Pair-wise comparisons from the interobserver study resulted in 1674 total concordant outcomes (761

![Figure 1](image_url)

**FIGURE 1.** Prevalence in non–small-cell lung cancer (NSCLC) tumor bank specimens stained with PD-L1 IHC 22C3. PD-L1 expression was detected over the entire dynamic range in NSCLC formalin-fixed paraffin-embedded specimens (A); using 50% tumor proportion score (TPS) as cut point, the percentage of clinical diagnostic positive cases was approximately 18.4% (B). Results were reported as the percentage of neoplastic cells showing membranous staining of PD-L1. Images shown are tumor samples obtained from patients with a TPS of <1%, a TPS of 1% to 49%, and a TPS of at least 50% (C).

### TABLE 1. Overall Response in PD-L1-positive Patients and in All Patients Treated With Pembrolizumab

| Response Evaluation | PD-L1 (TPS ≥ 50%, N = 61) | Total (N = 223) |
|---------------------|-----------------------------|-----------------|
|                     | n (%)                       | 95% CI          | n (%)                       | 95% CI          |
| Complete response (CR) | 0 (0.0)             | 0.0-5.9         | 0 (0.0)            | 0.0-1.6         |
| Partial response (PR)  | 25 (41.0)            | 28.6-54.3       | 46 (20.6)         | 15.5-26.5       |
| Overall response (CR + PR) | 25 (41.0)      | 28.6-54.3       | 46 (20.6)         | 15.5-26.5       |
| Stable disease (SD)   | 9 (14.8)              | 7.0-26.2        | 64 (28.7)         | 22.9-35.1       |

Only confirmed responses are included in this table.
Database cutoff date: August 29, 2014.
negative and 791 positive), with a total of 122 discordant outcomes. Performance criteria for ANA resulted in 92.6% with a lower bound of 87.8%. APA resulted in 92.8% with a lower bound of 88.1%, and OA resulted in 92.7% with a lower bound of 88.1% (Table 3). Acceptance criteria were met for all 3 agreement types; the lower bound of the 2-sided 95% CI was ≥ 85% for ANA, APA, and OA.

Pair-wise comparisons from the intraobserver study resulted in 558 total concordant outcomes (264 negative and 274 positive), with a total of 20 discordant outcomes. Performance criteria for ANA resulted in 96.4% with a lower bound of 94.0%. APA resulted in 96.5% with a lower bound of 94.3%, and OA resulted in 96.4% with a lower bound of 94.3% (Table 3). Acceptance criteria were met for all 3 agreement types; the lower bound of the 2-sided 95% CI was ≥ 85% for ANA, APA, and OA.

**DISCUSSION**

PD-L1 has been reported as a potential biomarker for assessment as a PD1/PD-L1 checkpoint inhibitor. Multiple IHC assays with different detecting antibodies, assay platforms, scoring methods, and cutoffs have been used with related clinical trials to detect PD-L1 expression as an indicator of clinical efficacy for several anti-PD1/PD-L1 therapeutics. A wide range of predictive response rates have been reported at the 2015 annual meeting of the American Society of Clinical Oncology. The Dako PD-L1 IHC 22C3 assay has been codeveloped with pembrolizumab to evaluate the correlation of PD-L1 expression with the clinical outcome for NSCLC patients in a clinical trial. Among all the patients tested, the objective response rate was 19.4%, and the median duration of response was 12.5 months. The median duration of progression-free survival was 3.7 months, and the median duration of overall survival was 12.0 months. When using 50% TPS as the cutoff for PD-L1 diagnostic positivity, the response rate increased to 45.2% and the median progression-free survival was prolonged to 6.3 months. 13

The Dako PD-L1 IHC 22C3 pharmDx assay has been developed to detect PD-L1 expression in FFPE human NSCLC tumor tissue specimens using Dako Autostainer Linker 48. Assessment of PD-L1 expression demonstrated staining across the dynamic range of 0% to 100% positivity and 0 to 3 staining intensity on tumor cells and immune cells. In a small study of tumor bank specimens of 127 unique cases, approximately 18% of the NSCLC patient specimens tested with PD-L1 IHC 22C3 pharmDx were identified as PD-L1 positive when using 50% TPS as cutoff. In other studies with larger sample size, results have shown that around 25% of NSCLC patients tested with this assay will be identified as PD-L1 positive. 13

Correct and accurate interpretation of PD-L1 staining by a pathologist ensures the effective use of PD-L1 inhibitors in the clinical setting. Although PD-L1
expression was seen on both tumor cells and immune cells, no predictive value was added by including immune cell staining in the KEYNOTE-001 clinical trial. Therefore only the PD-L1 expression on tumor cells is included in the interpretation. During PD-L1 IHC 22C3 assay development, a clear set of scoring guidelines and a comprehensive pathologist training program have been conducted and implemented in the clinical trial labs, which resulted to the high percentage agreements in the interpathologist repeatability and intrapathologist reproducibility of the PD-L1 IHC 22C3 pharmDx assay across multiple labs.

Internal and external studies have been performed to demonstrate the repeatability and reproducibility of the Dako PD-L1 IHC 22C3 pharmDx assay. The in-house analytical validation consisted of interinstrument, interoperator, interday, and interlot, and intraday and intrarun variations. NPA, PPA, and OA of 100% were achieved for all these studies. When the PD-L1 IHC 22C3 pharmDx assay reproducibility was tested at different external laboratories (sites), the concordance was slightly lower with an ANA of 90.8%, APA of 85.8%, and OA of 88.8%. This is to be expected since additional variables were added to the testing. Efforts were made to balance the number of positive specimens and negative specimens for the external reproducibility study. However, the final data showed that the study pathologists identified 21 diagnostic positive specimens, which contributed to the wider CI for APA (77% at 95% CI lower bound).

In conclusion, this study demonstrates that the Dako PD-L1 IHC 22C3 assay is a sensitive, specific, precise, and robust assay, which provides high value clinical utility to identify patients who will benefit from treatment with pembrolizumab.

REFERENCES
1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–674.
2. Sharpe AH, Freeman GJ. The B7-CD28 superfamily. Nat Rev Immunol. 2002;2:116–126.
3. Keir ME, Butte MJ, Freeman GJ, et al. PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol. 2008;26:677–704.
4. Freeman GJ, Long AJ, Iwai Y, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J Exp Med. 2000;192:1027–1034.
5. 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.
6. Gardon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med. 2015;372:2018–2028.
7. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature. 2014;515:563–567.
8. Tumeh PC, Harview CL, Yearly JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;515:568–571.
9. Mu C-Y, Huang J-A, Chen Y, et al. High expression of PD-L1 in lung cancer may contribute to poor prognosis and tumor cells immune escape through suppressing tumor infiltrating dendritic cells maturation. Med Oncol. 2011;28:682–688.
10. Gettinger S, Herbst RS. B7-H1/PD-1 blockade therapy in non-small cell lung cancer: current status and future direction. Cancer J. 2014;20:281–289.
11. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody MPDL3280A in cancer patients. Nature. 2014;515:568–571.
12. 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;366:2455–2465.
13. Boland JM, Kwon ED, Harrington SM, et al. Tumor B7-H1 and B7-H3 expression in squamous cell carcinoma of the lung. Clin Lung Cancer. 2013;14:157–163.
14. Chen YB, Mu CY, Huang J, et al. Clinical significance of programmed death-1 ligand-1 expression in patients with non-small cell lung cancer: a 5-year-follow-up study. Tumori. 2012;98:751–755.
15. Velchetti V, Schalper KA, Carvajal DE, et al. Programmed death ligand-1 expression in non-small cell lung cancer. Lab Invest. 2014;94:107–116.
16. Dolled-Filhart M, Roach C, Toland G, et al. Development of a companion diagnostic for pembrolizumab in non-small cell lung cancer using immunohistochemistry for PD-L1. Arch Pathol Lab Med. submitted.
17. Dolled-Filhart M, Roach C, Toland G, et al. Development of a PD-L1 immunohistochemistry (IHC) assay for use as a companion diagnostic for pembrolizumab (pembro; MK-3475) in non-small cell lung cancer (NSCLC). Proceedings: ASCO Annual Meeting; May 29–June 2, 2015; Chicago (abstract 11065). J Clin Oncol. 2015;33(suppl):11065.