

PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project



Fred R. Hirsch, MD, PhD,^{a,b,*} Abigail McElhinny, PhD,^c Dave Stanforth, MBA,^d James Ranger-Moore, PhD,^e Malinka Jansson, MA,^d Karina Kulangara, PhD,^d William Richardson, BA,^e Penny Towne, BS, MBA,^e Debra Hanks, MD,^d Bharathi Vennapusa, MD,^e Amita Mistry, MD,^e Rasika Kalamegham, PhD,^{f,g} Steve Averbuch, MD,^h James Novotny, PhD,^h Eric Rubin, MD,ⁱ Kenneth Emancipator, MD,ⁱ Ian McCaffery, PhD,^{j,k} J. Andrew Williams, PhD,^j Jill Walker, PhD,^l John Longshore, PhD,^m Ming Sound Tsao, MD,ⁿ Keith M. Kerr, MB, FRCPath^o

^aMedicine and Pathology, University of Colorado Cancer Center, Aurora, Colorado

^bInternational Association for the Study of Lung Cancer, Aurora, Colorado

^cPersonal Genome Diagnostics, Baltimore, Maryland

^dDako North America, Agilent Technologies, Carpinteria, California

^eVentana Medical Systems, Inc., Roche Tissue Diagnostics, Tucson, Arizona

^fAmerican Association for Cancer Research, Washington, DC

^gGenentech, Washington, DC

^hDevelopment, Oncology and Pharmacodiagnosics, Bristol-Myers Squibb Company, Princeton, New Jersey

ⁱMerck & Co., Inc., Kenilworth, New Jersey

^jGenentech, Oncology Biomarkers Development, South San Francisco, California

^kCorvus Pharmaceuticals, Burlingame, California

^lAstraZeneca, Cambridge, United Kingdom

^mCarolinas Pathology Group, Carolinas HealthCare System, Charlotte, North Carolina

ⁿDepartment of Pathology, University Health Network, Princess Margaret Cancer Centre and University of Toronto, Ontario, Canada

^oDepartment of Pathology, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen, Scotland

Received 14 November 2016; revised 23 November 2016; accepted 26 November 2016

Available online - 29 November 2016

*Corresponding author.

Drs. McElhinny and Stanforth equally contributed to this work.

Disclosure: Dr. Hirsch has received compensation from Genentech/Roche, Pfizer, Bristol-Myers Squibb, Lilly, Merck & Co., Inc., AstraZeneca, Boehringer-Ingelheim, and Ventana/Roche for participating in advisory boards and has received research funding (through the University of Colorado) from Genentech/Roche, Bristol-Myers Squibb, Lilly, Bayer, Amgen, and Ventana/Roche. Mr. Stanforth is a full-time employee of Dako and owns stock in Dako. Dr. Ranger-Moore is a full-time employee of and owns stock in Roche. Ms. Jansson is a full-time employee of and owns stock in Dako. Dr. Kulangara is a full-time employee of and owns stock in Dako. Mr. Richardson and his spouse are full-time employees of Roche. Ms. Towne is a full-time employee of Ventana Medical Systems, Roche Tissue Diagnostics, and owns stock in Roche Diagnostics. Dr. Hanks is a full-time employee of and owns stock in Dako. Dr. Vennapusa is a full-time employee of Ventana Medical Systems, Roche Tissue Diagnostics. Dr. Mistry is a full-time employee of Ventana Medical Systems, Inc., and owns stock in F. Hoffmann-La Roche Ltd. Dr. Kalamegham is a full-time employee of Genentech Inc. Dr. Averbuch is a full-time employee of Bristol-Myers Squibb. Dr. Novotny is a full-time employee of Bristol-Myers Squibb. Dr. Rubin is a full-time employee of Merck & Co., Inc. Dr. Emancipator is a full-time employee of Merck & Co., Inc. and owns stock in Merck & Co., Inc., Bayer AG, and Johnson and Johnson. Dr. McCaffery's

spouse was an employee and stockholder of Genentech/Roche during the writing of this study. Dr. Williams is an employee of Genentech. Dr. Walker is a full-time employee of and owns stock in AstraZeneca. Dr. Longshore has performed contract research for Agilent Technologies and Ventana; has received compensation from Ventana, AstraZeneca, Bristol-Myers Squibb, and Genentech for participating in advisory boards; and has received consultancy fees and/or honoraria from Ventana, AstraZeneca, Bristol-Myers Squibb, Genentech, and Merck & Co., Inc. Dr. Tsao has received compensation from AstraZeneca, Merck & Co., Inc., Ventana/Roche, and Bristol-Myers Squibb for participating in advisory boards and has received research funding (through the University Health Network) from Merck & Co., Inc., Canada. Mr. Kerr has received consultancy fees and/or honoraria from Roche/Genentech, AstraZeneca, Bristol-Myers Squibb, Merck & Co., Inc., and Ventana. The remaining author declares no conflict of interest.

Address for correspondence: Fred R. Hirsch, MD, PhD, University of Colorado Cancer Center, 12801 E. 17th Ave., MS:8117, Building RC1 South, Room 8119, Aurora, CO 80045. E-mail: Fred.Hirsch@ucdenver.edu

© 2017 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

ISSN: 1556-0864

<http://dx.doi.org/10.1016/j.jtho.2016.11.2228>

ABSTRACT

Introduction: The Blueprint Programmed Death Ligand 1 (PD-L1) Immunohistochemistry (IHC) Assay Comparison Project is an industrial-academic collaborative partnership to provide information on the analytical and clinical comparability of four PD-L1 IHC assays used in clinical trials.

Methods: A total of 39 NSCLC tumors were stained with four PD-L1 IHC assays (22C3, 28-8, SP142, and SP263), as used in the clinical trials. Three experts in interpreting their respective assays independently evaluated the percentages of tumor and immune cells staining positive at any intensity. Clinical diagnostic performance was assessed through comparisons of patient classification above and below a selected expression cutoff and by agreement using various combinations of assays and cutoffs.

Results: Analytical comparison demonstrated that the percentage of PD-L1-stained tumor cells was comparable when the 22C3, 28-8, and SP263 assays were used, whereas the SP142 assay exhibited fewer stained tumor cells overall. The variability of immune cell staining across the four assays appears to be higher than for tumor cell staining. Of the 38 cases, 19 (50.0%) were classified above and five (13%) were classified below the selected cutoffs of all assays. For 14 of the 38 cases (37%), a different PD-L1 classification would be made depending on which assay/scoring system was used.

Conclusions: The Blueprint PD-L1 IHC Assay Comparison Project revealed that three of the four assays were closely aligned on tumor cell staining whereas the fourth showed consistently fewer tumor cells stained. All of the assays demonstrated immune cell staining, but with greater variability than with tumor cell staining. By comparing assays and cutoffs, the study indicated that despite similar analytical performance of PD-L1 expression for three assays, interchanging assays and cutoffs would lead to “misclassification” of PD-L1 status for some patients. More data are required to inform on the use of alternative staining assays upon which to read different specific therapy-related PD-L1 cutoffs.

© 2017 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

Keywords: Immunotherapy; Lung cancer; PD-L1 assays; Immunohistochemistry

Introduction

Immunotherapies with checkpoint inhibitor programmed death ligand 1 (PD-L1)/programmed cell death 1 antibodies have shown encouraging results in patients with advanced NSCLC.¹⁻¹⁰ Three agents, pembrolizumab (Keytruda [Merck & Co., Inc., Kenilworth, NJ]), nivolumab

(Opdivo [Bristol-Myers Squibb, Lawrenceville, NJ]), and atezolizumab (Tecentriq [Genentech/Roche, South San Francisco, CA]) are approved by the U.S. Food and Drug Administration (FDA) for patients with advanced NSCLC after failure of first-line therapy, whereas durvalumab (AstraZeneca, Wilmington, DE) is still under clinical development for use in NSCLC. Furthermore, pembrolizumab was recently approved by the FDA for first-line therapy for patients with advanced NSCLC.¹¹ Clinical trials with each of these drugs have shown an association between the magnitude of clinical efficacy and level of tumoral PD-L1-expression as evaluated by PD-L1 immunohistochemistry (IHC).^{1-5,7-10} In 2015, the FDA approved two PD-L1 IHC assays along with their corresponding drugs for NSCLC—one using the 22C3 clone as a “companion diagnostic” for pembrolizumab and one using the 28-8 clone as a “complementary diagnostic” associated with nivolumab.^{12,13} Most recently, a complementary diagnostic was approved for atezolizumab.¹⁴

The current one drug—one diagnostic test codevelopment approach for approval of therapeutic products in stratified or selected patient populations has resulted in each of the four therapeutic agents that are either FDA-approved or in late-stage development being associated with a unique anti-PD-L1 IHC assay. Two of the approved NSCLC assays are manufactured by Dako (Carpenteria, CA) and are optimized for use with the detection systems developed for the Dako Link 48 staining platform.^{15,16} The other two assays (one approved and one not yet approved) have been developed with different detection technologies on the Ventana BenchMark platform (see Table 1). Each IHC assay was developed with a unique primary antibody (clone) against PD-L1, namely, 28-8 (Dako) with nivolumab [Bristol-Myers Squibb], 22C3 (Dako) with pembrolizumab [Merck & Co., Inc.], SP263 (Ventana) with durvalumab [AstraZeneca], and SP142 (Ventana) with atezolizumab [Genentech]. Critically, the clinical scoring approaches for each of the four diagnostic assays used to classify patients for treatment on the basis of tumoral PD-L1 expression utilize a measure of PD-L1 expression on tumor cell (TC) membranes. In addition to TCs, Genentech/Roche also includes a measure of infiltrating PD-L1-positive immune cells (ICs) as part of the scoring guideline the SP142 assay for atezolizumab (see Table 1).^{9,10} These scoring approaches were determined on the basis of predictive value and clinical data from individual drug development programs during the therapeutic/diagnostic test codevelopment process, resulting in individual PD-L1 diagnostic systems uniquely tailored for each immune checkpoint inhibitor.

The approval of multiple PD-L1 IHC assays to identify appropriate therapies within a single class poses a unique challenge to patients and care providers with

Table 1. PD-L1 Assay Systems Used in the Blueprint Project				
Agent	Nivolumab	Pembrolizumab	Atezolizumab	Durvalumab
Primary antibody clone used in the assay system	28-8 (Dako)	22C3 (Dako)	SP142 (Ventana)	SP263 (Ventana)
Interpretive scoring	Tumor cell membrane	Tumor cell membrane	Tumor cell membrane Infiltrating immune cells	Tumor cell membrane
Instrument and detection systems required	EnVision Flex on AutostainerLink 48	EnVision Flex on AutostainerLink 48	OptiView detection and amplification on Benchmark ULTRA	OptiView detection on Benchmark ULTRA
Therapeutic developer	Bristol-Myers Squibb	Merck	Genentech	AstraZeneca

PD-L1, programmed death ligand 1.

respect to clinical application of PD-L1 testing and treatment decision making.¹⁷⁻¹⁹ The limited availability of tumor tissue for testing, the number of tissue-based diagnostic tests required in the management of a patient with NSCLC, and the complexity of testing and interpretation with multiple tests for this same PD-L1 analyte drive a need to understand the correlation of results across the PD-L1 tests available. The medical community requires clarity on how to use these different PD-L1 diagnostic IHC assay systems to drive efficiency and safeguard the integrity of treatment decision making, especially when faced with calls for one PD-L1 test for all drugs. This medical community includes patients, pathologists, oncologists, health authorities, and payers.

To enable a better understanding of similarities and differences between these four PD-L1 IHC systems, the Blueprint PD-L1 IHC Assay Comparison Project was founded.²⁰ This project is a collaboration between the International Association for the Study of Lung Cancer and the American Association for Cancer Research, together with four pharmaceutical companies (Bristol-Myers Squibb, Merck & Co., Inc., AstraZeneca, and Genentech/Roche) and two diagnostic companies (Dako/Agilent and Ventana/Roche). A key focus of the study was an assessment of the analytical similarities and differences between the PD-L1 systems to better understand their technical performance.

The Blueprint PD-L1 IHC Assay Comparison Project is planned in two phases. The result of phase 1, presented here, is a feasibility assessment that included a limited number (n = 39) of NSCLC samples stained with all four investigational use-only assays, as used in clinical trials and assessed by trained pathologists from the two diagnostic companies (Ventana/Roche and Dako/Agilent). The goal of phase 1 was to compare the analytical staining factors reported as percentages of stained cells, as well as selected treatment-determining scoring algorithms developed for each assay and used in clinical trials. Testing of clinical trial samples and correlation to clinical outcome were beyond the scope of this study. The results of this study cannot determine whether any of assays are more specific and/or sensitive or better or

worse for treatment decision making. The results from phase 1 will be used to design a larger, more comprehensive phase 2 study.

Materials and Methods

NSCLC Sample Cohort

Human formalin-fixed, paraffin-embedded NSCLC samples were obtained independently by Dako and Ventana from third-party vendors. These samples were not associated with any therapeutic trial or immune checkpoint inhibitor therapy. A total of 39 specimens (18 obtained by Dako and 21 by Ventana, the vast majority of which were from surgical resections and a few of which were from biopsies) were selected to represent the respective dynamic range of each of the four PD-L1 assays. The samples demonstrated a range of TC staining, IC staining, and combined TC and IC staining. The samples' distribution did not represent the distribution of PD-L1 expression reported in trials, and the samples were not preselected by histological subtype, stage, sex, or other demographic criteria. No therapeutic outcome data were available. The blocks were sectioned at 4 μm at Dako and Ventana, and at least 10 sections were serially cut from each block for hematoxylin and eosin and IHC staining. The cut slides were exchanged well within the cut slide stability time limit for each assay (4–6 months) for staining by the respective company laboratories (Fig. 1).

PD-L1 IHC Assays

Dako and Ventana independently stained each of the 39 cases using their respective PD-L1 IHC assay platforms. At Dako, sections were stained with anti-PD-L1 28-8 rabbit monoclonal primary antibody or anti-PD-L1 22C3 mouse monoclonal primary antibody by utilizing the EnVision FLEX visualization system on a Dako Autostainer Link 48 system with negative control reagents and cell line run controls as described in the PD-L1 IHC 22C3 pharmDx and PD-L1 IHC 28-8 pharmDx package inserts.^{21,22}

At Ventana, sections were stained with anti-PD-L1 (SP263) rabbit monoclonal primary antibody²³ and a

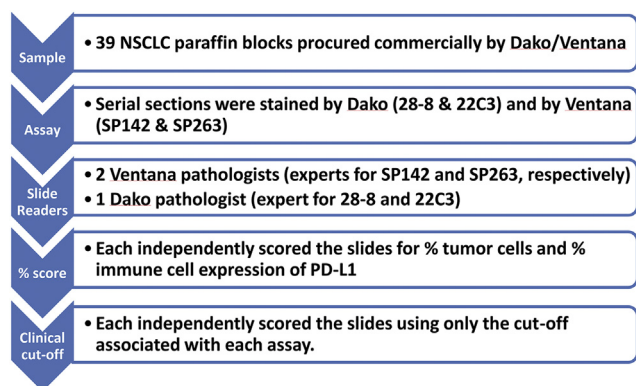


Figure 1. Blueprint Project: materials and methods workflow.

matched rabbit immunoglobulin G–negative control with an OptiView DAB IHC Detection Kit on the BenchMark ULTRA automated staining platform. For the SP142 assay, sections were stained with anti-PD-L1 (SP142) rabbit monoclonal primary antibody and a matched rabbit immunoglobulin G–negative control with an OptiView DAB IHC Detection Kit followed by an OptiView Amplification Kit, also on the BenchMark ULTRA automated staining platform.⁹

In addition to the PD-L1 IHC staining, hematoxylin and eosin staining was performed for each case to help orientate the pathologists' reading of the IHC slides.

Analytical Performance Comparison

Three pathologists, all of whom were experts in interpreting their respective clinical cutoffs of the assays used in this study, independently evaluated all 156 immunostained slides from the 39 cases, making a total of 468 observations of raw percentages of cells expressing PD-L1 without consideration of any cutoff level. (Pathologist 1 expert in scoring the 22C3 and 28-8 assays, pathologist 2 in the SP142 assay, and pathologist 3 in the SP263 assay). The pathologists did not discuss any of the staining interpretations beforehand, nor did they confer during the scoring process. There was no discrepancy review for discordant results. Consequently, when evaluating analytical staining components reported as percentages, the pathologists were not blinded to the specific assay.

The following analytical components were assessed: TC staining (an estimate of the percentage of TCs exhibiting partial or complete membranous staining) and IC staining (an estimate of the percentage of ICs, including macrophages and lymphocytes within the tumor, exhibiting staining). The assessment of the percentage of IC-infiltrated tumor region was not part of this particular initial analysis (but see later). ICs were identified by morphologic features only, without the use

of other IC biomarkers. A semiquantitative scoring system (estimation) was used to calculate the percentage of TCs exhibiting membrane staining (tumor proportion score [TPS]), and the percentage of ICs stained (membrane or cytoplasm) for each slide was assessed. Negative reagent controls were evaluated in each case to confirm acceptable background staining. Staining intensity was not part of the evaluation.

Clinical Diagnostic Performance Comparison

Clinical performance comparison refers to the evaluation of cell staining using a scoring algorithm and pre-specified cutoff, as would be used for treatment decision making. No actual patient clinical outcome data are available for these 39 cases. For the clinical diagnostic assessment, one case was assessed by the pathologists as containing insufficient material; this case was eliminated from further analysis, leaving 38 samples for comparison. The 152 slides (38×4) from the 38 samples were randomized and blinded to specific PD-L1 IHC assay, and the three pathologists independently evaluated each stained sample according to the preselected cutoff chosen for each assay. Although several cutoffs relating to levels of PD-L1 expression have been used in clinical trials, for this study the companies submitted a single clinical cutoff for inclusion in this comparative analysis; 1% TC staining for the 28-8 and 22C3 assays, 25% TC staining for the SP 263 assay, and 1% TC staining and/or 1% tumor area infiltrated by PD-L1–positive ICs (TC1/IC1) for the SP142 assay (see [Table 1](#)).

Each expert pathologist scored all 152 slides by using only the scoring algorithm(s) with which they had greatest expertise (thus, pathologist 1 scored all 152 slides, applying the 22C3 and 28-8 algorithms; pathologist 2 scored using the SP142 algorithm; and pathologist 3 scored according to the SP263 algorithm). All results from these reads were reported as being either above or below the respective selected clinical cutoff.

Data Analysis

Data were analyzed and reported with respect to agreement between (1) the estimated percentage of expression status ([Figs. 2 and 3](#)) and (2) the case status with respect to the selected clinical algorithm (i.e., above or below the assay cutoff) for each assay, in the 38-case cohort ([Fig. 4](#)). For the calculation of overall agreement, the score of the pathologist who was an expert on the given assay was used as a reference. For example, pathologist 1 score was used as the reference standard for 22C3-stained slides, pathologist 3 score was used for the SP263-stained slides, and so forth.

Separate analyses were conducted by interchanging various cutoffs on each set of slides stained by the four different assays ([Table 2](#)) and then comparing the

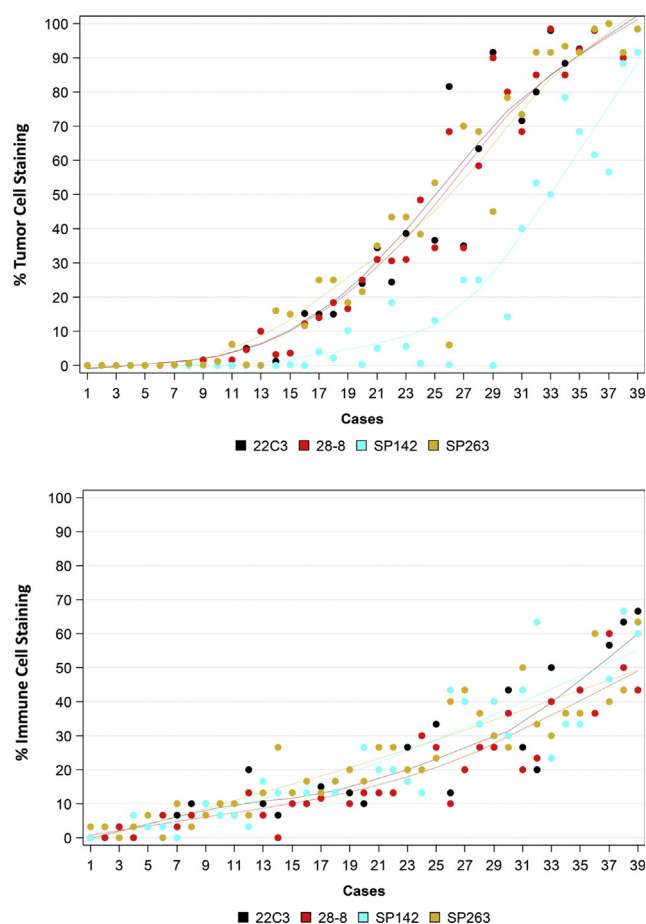


Figure 2. Analytical comparison of percentage tumor cell and immune cell staining, by case, for each assay. Data points represent the mean score from three pathologists for each assay on each case. Superimposed points indicate identical values. No clinical diagnostic cut-off was applied. ‘Best fit’ colored curves allow comparison of score range between the four assays.

overall agreement for each combination with the index scores derived according to the assay and cutoff combination selected and specified for the purposes of this study for each drug and assay combination, herein referred to as the *validated assay and cutoff combination*. Thus, the assessment by pathologist 1 using the 28-8 algorithm on the SP142-, SP263-, and 22C3-stained slides was used to determine what proportion of cases would differ when compared with the cutoff that was developed for the 28-8 assay. Other combinations of assays and cutoffs were also used in this analysis as detailed in Table 2.

Analyses and Reporting of Analytical Performance

Frequency distribution graphs for the percentage of cells stained for both TCs and ICs were generated for each assay as evaluated by each pathologist (see Fig. 2).

Data represent the mean derived from each pathologist’s results as well as fitted lines showing the general relationship between assays. Scatter plots were also constructed to show pairwise comparisons for the percentage of cell staining for both TCs and ICs (see Fig. 3) for individual pathologists, as well as regression analysis for correlation between assays and pathologists. Analyses were conducted independently by statisticians at Dako and Ventana, and any discrepancies were resolved, resulting in a single final quality-controlled set of results.

Statistical software products used were SAS version 9.4 (SAS Institute Inc., Cary, NC) and Stata/SE version 12.1 (StataCorp, College Station, TX) at Ventana and R version 3.2.0 (R Foundation for Statistical Computing, Vienna, Austria) at Dako.

Analyses and Reporting of Clinical Performance

To assess the agreement of the four PD-L1 IHC assays (scoring using the algorithm selected for each specific staining assay—the validated assay and cutoff combination) on the 38-case cohort, a heat map and a Venn diagram were generated (see Fig. 4) to compare the staining results aligned by case for each assay and arranged by increasing level of PD-L1 staining observed.

Analyses at the clinical assay level (see Table 2) were also carried out by looking at overall agreement when comparing the trial-validated assay and cutoff combination as a reference with the alternative algorithm and cutoff combinations applied to the same assay.

Results

Analytical Performance Comparison

PD-L1 Staining of TCs and ICs. PD-L1 staining was observed in both TCs (Fig. 5A–D and F) and ICs (Fig. 5A–E and G) by all four of the PD-L1 assays: 22C3, 28-8, SP263, and SP142.

TC staining by 22C3, 28-8, and SP263 showed a range of intensities and partial or full circumferential membrane staining (intensity not included in the comparative analysis, only percentage of stained cells). Overall, these assays showed relative staining equivalency in TCs. In most cases, SP142 showed weaker staining of TC membranes and fewer positive TCs compared with the other three assays (see Fig. 5B–D). Figures 5A to G depict seven NSCLC samples that represent a range of intensities for PD-L1 with all four assays. Figure 5A represents high TC staining intensity with all four assays, whereas Figure 5E and G shows no TC staining with any of the four assays. Figure 5D depicts weak to moderate intensity of TC staining when the four assays are compared. Figures 5B and C demonstrate weak tumor staining intensity.

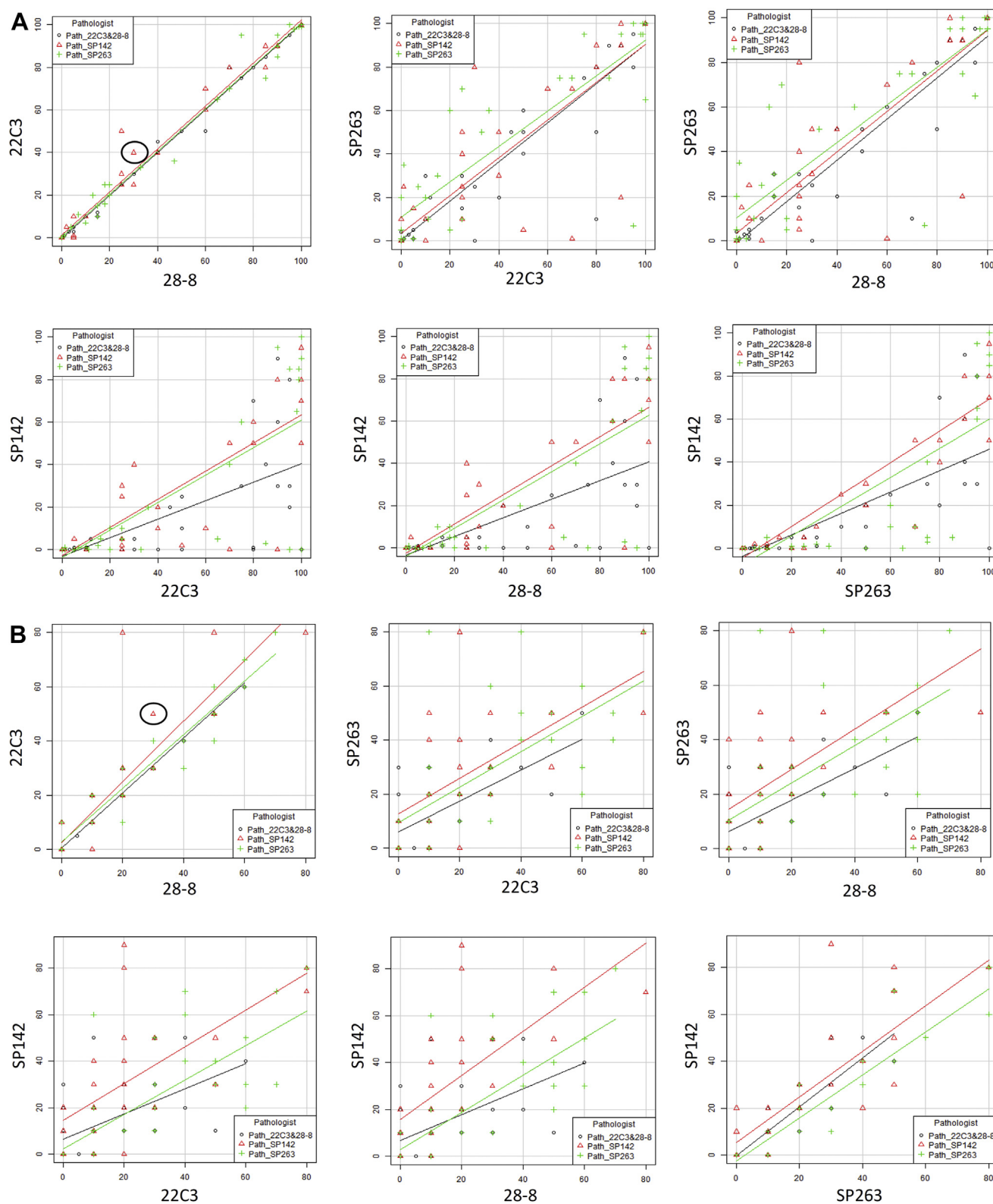


Figure 3. Pairwise associations for percentage tumor or immune cell staining presented by reader. Each scatter plot shows the pair-wise relationship of percentage tumor (A) or immune (B) cell scoring on 38 cases by three independent observers. The data is presented by assay pairs in each plot such that the top-left scatter plot illustrates the pair-wise data (on a scale of 1-100%) between assays 22C3 and 28-8. The differently-shaped colored symbols represent data of individual readers on the 38 cases, and the corresponding colored regression lines show the correlation between each assay pair based on the scores for each reader. For 4A as an example, the circled red triangle in the upper left plot is for a case that was scored at 40% tumor cells positive with 22C3 and 30% tumor cells positive with 28-8 by observer Path_SP142. Some points are superimposed on the plots but there are 38 data points per observer in all plots. A 45-degree regression line indicates perfect correlation and superimposed regression lines indicate low inter-observer variability.

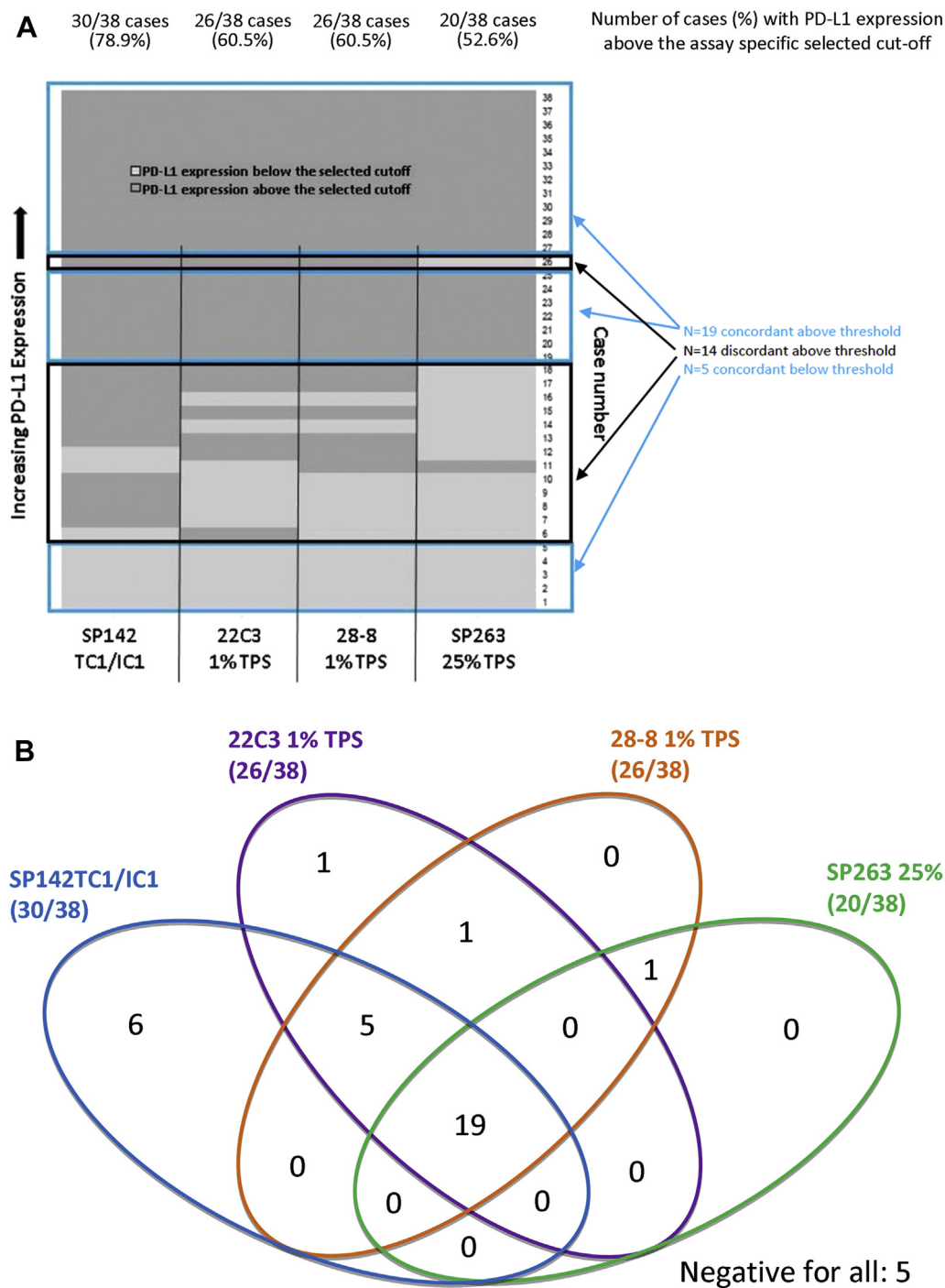


Figure 4. Comparison of cases allocated above or below clinical assay thresholds. Scoring applied according to threshold for the assay. Heat map (A) and Venn diagram (B) show the diagnostic PD-L1 classification of the 38 cases stained with each of the four PD-L1 IHC systems and scored above and below their respective clinical threshold. The map in (A) illustrates the diagnostic outcome (concordance/discordance) for each case (rows 1–38 shown on the right hand side) across the various assays/scoring algorithm combinations. The light gray color corresponds to PD-L1 levels below each respective threshold while the dark gray color corresponds to PD-L1 levels above each respective threshold. Five cases out of the total 38 show concordance below all threshold values across all assay/algorithm combinations, while 19 cases show expression above all threshold values across all assay/algorithm combinations. The remaining 14 cases show a combination of discordant outcome across the various assay/algorithm combinations.

IC staining was observed with all four assays, thus detecting infiltrating ICs. Representative images are shown in Figures 5A–E and G. Figures 5A, C, and F depict low IC staining with all four PD-L1 assays. Figures 5D and E show moderate IC staining with the four assays whereas Figure 5B depicts high IC staining with all four PD-L1 assays.

Table 2. Assay Comparison: Overall Percentage of Agreement in Patient Classification When Staining Assays Are “Mismatched” with the Clinical Cutoff

Assay Clone Used for Slide Staining	Scoring Algorithm							
	22C3	1% TPS	28-8	1% TPS	SP142	TC1/IC1	SP263	25% TPS
22C3	38 of 38	(100%)	36 of 38	(94.7%)	33 of 38	(86.8%)	34 of 38	(89.5%)
28-8	36 of 38	(94.7%)	38 of 38	(100%)	31 of 38	(81.6%)	33 of 38	(86.8%)
SP142	24 of 38	(63.2%)	24 of 38	(63.2%)	38 of 38	(100%)	25 of 38	(65.8%)
SP263	34 of 38	(89.5%)	34 of 38	(89.5%)	33 of 38	(86.8%)	38 of 38	(100%)

Note: Table indicates the number of cases that were concordant with the index assay scoring algorithm (assay and matching scoring algorithm) when an alternative cutoff was used to determine the allocation of cases to clinical groups above and below the cutpoint. TPS, tumor proportion score.

SP142 typically displayed more punctate, discontinuous staining, which is a reflection of the amplification components used in the detection system for the SP142 assay.

The distribution of TC staining of any intensity for each assay, pooled across pathologists, is shown in Figure 2, which shows the mean PD-L1 expression as assessed by the three pathologists plotted as a percentage of TC staining. Three of the assays (28-8, 22C3, and SP-263) had very similar distribution across the cases, whereas one assay (SP-142) consistently showed fewer TCs expressing PD-L1.

The distribution of IC staining of any intensity for each assay pooled across pathologists is also shown in Figure 2. The variability in staining pattern between cases was more pronounced for ICs than for TCs.

Pairwise Comparisons of PD-L1 Staining of TCs and ICs. The percentage of PD-L1-positive TCs stained, as evaluated by each pathologist for all possible pairwise comparisons between the four assays, is shown in Figure 3A, in which Path_22C3 and 28-8 indicates the expert pathologist in both assays (pathologist 1), Path_SP142 indicates the expert pathologist in the SP142 assay (pathologist 2), and Path_SP263 indicates the expert pathologist in the SP263 assay (pathologist 3). A 45-degree regression line indicates perfect correlation. The three lines indicate regression lines for each observer. Each plot is a comparison of the assays indicated in the x and y axis labels.

The 22C3, 28-8, and SP263 assays demonstrated a high correlation for numbers of stained TCs, as shown by the nearly 45-degree regression lines indicating linear relationships between assay pairs (Fig. 3A). Among those three assays, 22C3 versus 28-8 showed minimal inter-assay variability (least scattered points) and the highest score correlation (most superimposed lines) relative to 22C3 versus SP263 and 28-8 versus SP263. It should be noted that 22C3 and 28-8 were read by the same reader, so the latter two comparisons may add a component of interreader variability compared with the comparison of 22C3 versus 28-8.

All comparisons that include SP142 show lower correlation between assays (<45-degree regression lines) and more variability between assessments (lines not superimposed), indicating lower levels of positive TCs when stained with SP142.

Figure 3B depicts assessment of infiltrating IC staining by pathologist for all possible pairwise comparisons between the four assays. The slopes of the regression diagonal lines were less than 45 degrees for most pairwise comparisons, indicating that the assay on the x axis of the graph had fewer IC staining than the assay associated with the y axis. The 22C3 and 28-8 assays, however, demonstrated almost identical results. In all comparisons, the variability of IC staining was greater than the variability of TC staining. There is also greater variation in assay read with respect to ICs, as is evidenced by the fact that there is practically no overlap of the regression lines.

Clinical Diagnostic Performance Comparison

PD-L1 Expression as Defined by the Four Selected and Validated PD-L1 Assays. The PD-L1 classifications (above or below the assay threshold) for 38 of the 39 NSCLC samples were compared to determine concordance between the four assays when classifying cases according to the assay and cutoff combination selected and specified for the purposes of this study for each drug and assay combination, herein referred to as the *validated assay and cutoff combination*. Each of the expert pathologists' evaluations for their specific assay (e.g., SP263 slides evaluated by the SP263 pathologist, 22C3 slides evaluated by the 22C3 pathologist, etc.) was used in this comparison.

The heat map in Figure 4 illustrates, on a case-by-case basis, those cases with tumors expressing PD-L1 at levels above or below the validated cutoff for each assay. The light gray shows cases that the expert pathologist for each assay assessed as being below the validated PD-L1 cutoff value, whereas the dark gray shows cases that were at or above the validated cutoff value. Both light and dark for a particular case indicates that the case was evaluated differently depending on the

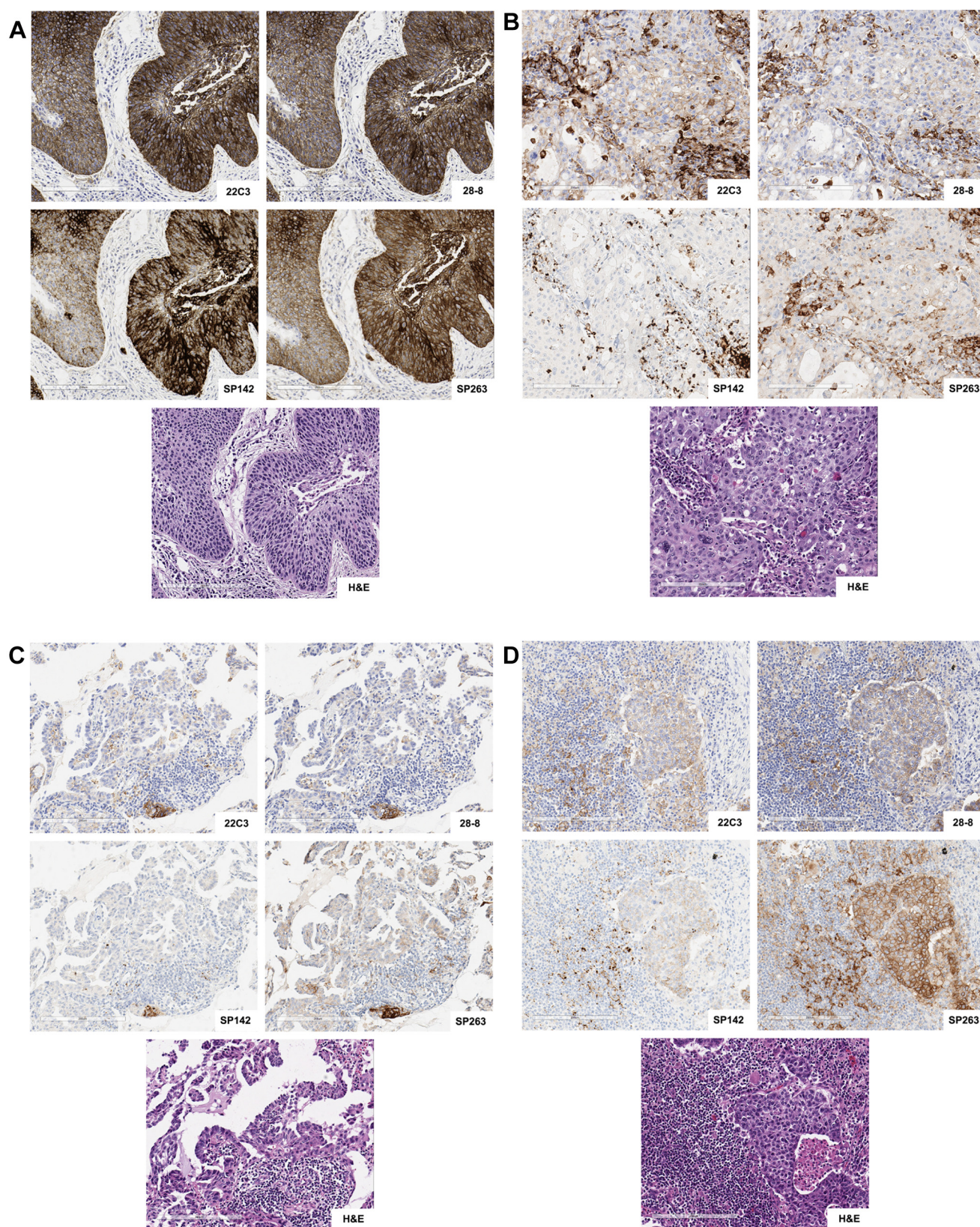


Figure 5. Representative staining of tumor cells from weak or negative to strong staining on five representative NSCLC cases by the four PDL1 assays at 20X magnification (A, B, C, D, E, and F) and 10X (G). 22C3, 28-8 and SP263 assays appeared relatively similar in overall pattern and coverage of tumor cells (brown), while SP142 stained fewer tumor cells in the majority of cases. In A, B, C, D, E, and G representative staining of infiltrating immune cells are shown across a range of expression. Note that all four assays detected immune cells to varying degrees.

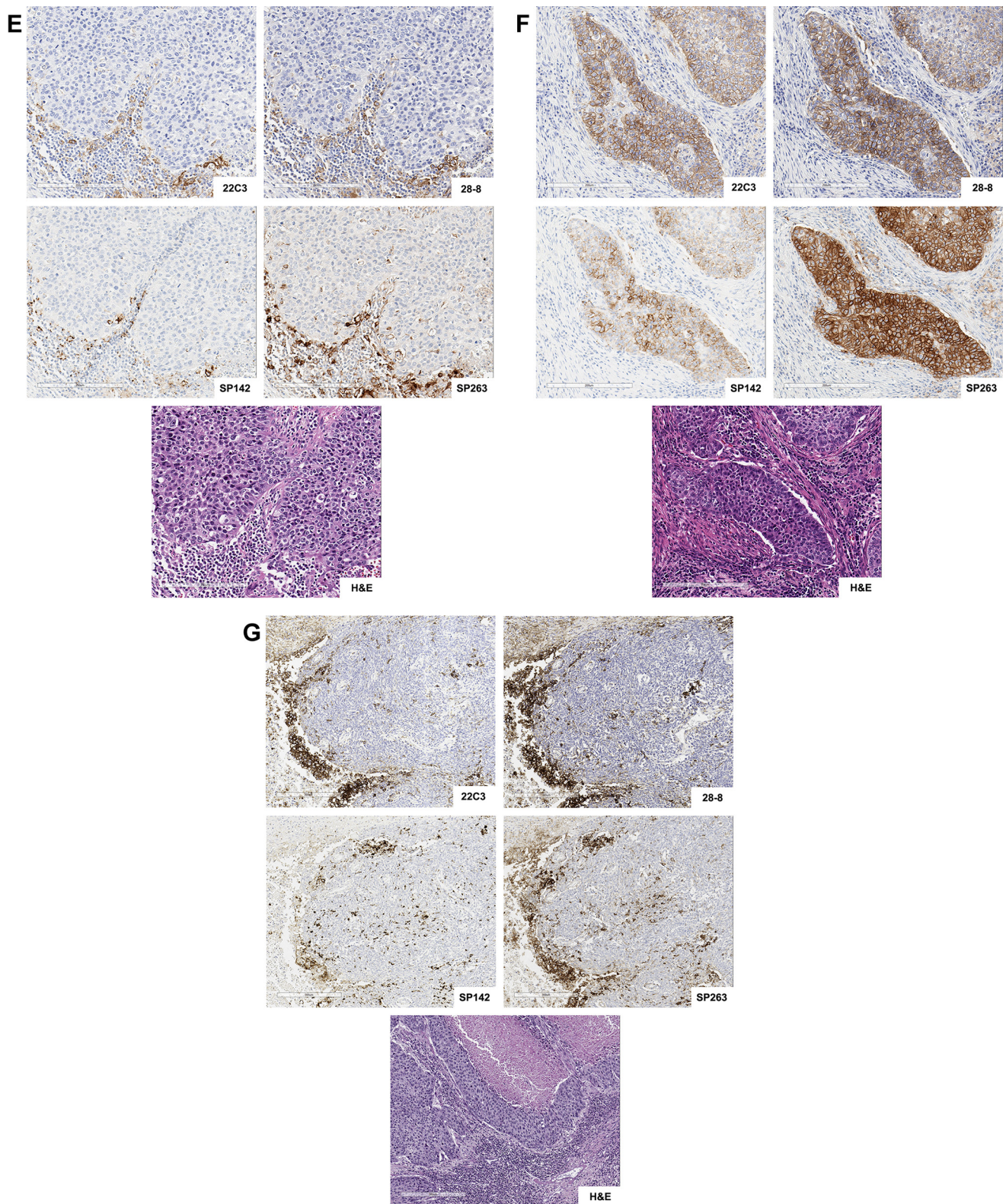


Figure 5. (continued).

validated assay and cutoff used. The prevalence of cases above the threshold decreases from the left of the figure to the right: the SP 142 assay showed 30 of 38 cases (78.9%) above the cutoff value, with the 22C3 assay

showing 26 of 38 cases (60.5%), the 28-8 assay showing 26 of 38 cases (60.5%), and the SP263 assay showing 20 of 38 cases (52.6%) above the selected cutoff value, respectively.

Nineteen of the 38 cases (50.0%) were above the cutoffs utilized by all four assays, meaning that clinical PD-L1 positivity would be concordant regardless of the assay used. Fourteen cases (37%) showed discordance between clinical levels of PD-L1 expression. Five of 38 (13%) samples were determined to be below the cutoff regardless of the assay used. These data indicate that in these 38 cases, use of an alternative validated assay and the assay-associated scoring algorithm to evaluate PD-L1 expression would give different results in terms of classifying a case above or below a treatment-determining threshold for a chosen therapy in slightly more than a third of cases.

Clinical Performance When Assays and Cutoffs Are Interchanged. Table 2 shows how the overall percentage of agreement changes for the 38 cases when, for each particular staining assay, each of the four cutoffs (scoring algorithms) for each individual assay is applied. In each comparison, the cutoff is held constant and the assay is interchanged. The reference standard for each comparison is the validated cutoff and assay combination. For instance, the application of the SP142 TC1/IC1 scoring algorithm to the SP142 stained slide results in a 100% rate of agreement by definition; however, applying the SP263 algorithm (which is positive at $\geq 25\%$ TPS) to SP142 slides results in a 65.8% rate of agreement.

In all situations, replacement of the validated cutoff for each assay with any other cutoff reduces the overall agreement compared with the reference standard.

Discussion

The data generated in this feasibility study provide early insights into relative analytic comparisons of the four trial-validated PD-L1 IHC assays in NSCLC. Since the study began, three of the PD-L1 assays have been approved by the FDA for use in NSCLC, and one is expected to gain approval for this indication in the near future. This unique collaboration was established to include both the relevant pharmaceutical stakeholders as well as the diagnostic companies and the independent academic organizations International Association for the Study of Lung Cancer and American Association for Cancer Research. This is the first and only study that has compared these four investigational use PDL-1 IHC assays from clinical trials pursued by the different pharmaceutical companies as companion diagnostics or complementary diagnostics.

Three key elements apply to this comparative study: the use of consecutive sections from the same tissue samples for testing by all four PD-L1 IHC assays, the application of each assay as prescribed by its use in

clinical trials (instrument platform, staining protocol, and scoring guidelines), and the reading of the individual assays by experts on those assays (a definitive standard assessment). There were essentially four comparisons made of the data generated by the pathologists: (1) comparison on a case-by-case basis of the percentage of tumor and ICs stained by each assay, as assessed by each pathologist across the range of possible scores (0%–100%) (Fig. 2); (2) a pairwise comparison of the different assays, as assessed by each pathologist (Fig. 3); (3) a comparison of clinical diagnostic performance in which each assay was read by following the algorithm (preselected cutoff) defined in trials for that specific assay (Fig. 4A and B); and (4) a comparison of clinical assay performance in which each assay was read according to alternative scoring algorithms (alternative definitions of cutoff) selected for the other three assays in the study (see Table 2).

This current work is a feasibility study, and the numbers of NSCLC samples and pathologists are small. Nonetheless, these data will begin to inform questions already being discussed in the wider community. How interchangeable might these assays be? Can an alternative assay be used, and if so, should it be read according to the rules for the assay used or for the drug to be prescribed? It should also be emphasized that the Blueprint Project is purely a study comparing assays' technical performance; there are no data available on the clinical predictive power of alternative PD-L1 IHC testing strategies.

Although a strength of the current study is that "experts" in staining and interpretation were used in the assay comparisons, the authors acknowledge that this feasibility study does not reflect the real-world situation, as the slides were read by experts in a particular assay and the generality of the results to non-company experts is unknown. Furthermore, the vast majority of specimens were surgically resected specimens and not biopsy specimens (the most frequent type of diagnostic specimens from advanced NSCLC). A comparison of surgically resected specimens (large specimens) and smaller biopsy specimens is planned in phase 2 of the Blueprint Project. A recent published study by Ilie et al. reported an eventual difference in PD-L1 expression when biopsy specimens and surgically resected specimens are compared.²⁴

The results of this feasibility study indicate that there are both similarities and differences with respect to the four PD-L1 systems in terms of dynamic ranges, cell types stained, and overall staining characteristics. Overall, three of the assays (28-8, 22C3, and SP263) were similar in analytical staining performance assessed by percentage of tumor cells showing cell membrane staining. The SP142 assay generally stained fewer TCs

across the 39 cases, which was also reported recently in a German comparison study by Scheel et al.²⁵ However, this assessment cannot differentiate between greater sensitivity for 28-8, 22C3, and SP263; greater specificity for SP142; or both. For IC staining, all four assays detected ICs but to a different extent, which was probably confounded by pathologist variability in the overall scores for ICs across the assays. Variability in IC scoring is likely due to the fact that the pathologists did not previously train or align on criteria for scoring the IC components, whereas scoring of TC membrane staining is more routinely performed for IHC, is more standardized, and represents a skill more transferable from experience of alternative assays. IC staining can be approached in different ways and is not a routine clinical practice. Also, little is known about possible differential staining by these assays of different IC types. The suggestion of greater variance in the IC scoring suggests that this unfamiliar practice will pose some challenges when first attempted, although training and practice in IHC reading skills are known to improve reproducibility.²⁶ The tendency for expert pathologists, using their specific algorithms, to score differently from the untrained pathologists emphasizes that training on the scoring algorithms for these assays will be critical for reproducibility among pathologists. This is especially true, as supported by our limited data, for the assessment of ICs. Related to assessment of ICs, it should be emphasized that for the clinical diagnostic algorithm for SP 142 used in clinical trials (TC and/or IC assessment) and in our clinical diagnostic comparisons (and described in the manual), IC expression is based on the percentage of tumor area occupied by IC staining and not as used in our analytical comparison (percentage of ICs). It should also be noted that this study cannot adequately address the issue of reproducibility, given that it involved only a limited number of experts who read the slides. This, however, gives a strong baseline for comparing assay performance that will not be confounded by a lack of pathologist training and experience.

When pathologists applied the selected algorithm and/or cutoff appropriate for each assay to determine the PD-L1 expression status of each case (the matched algorithm analyses shown in the heat map in Fig. 4), 50% of cases were above the respective prespecified threshold for each assay. A further 13.1% were unanimously below all of the selected cutoffs whereas slightly more than one-third of cases (36.9%) varied in classification above or below the assay-associated cutoff. Given the underlying similarity in actual staining similarities as assessed by raw percentage of positive staining TCs and ICs, as demonstrated in Figure 2, the differences observed in this analysis are likely due to the variation in definition of cutoff. Such are the differences in selected

cutoffs between validated clinical assays, and it was inevitable that there would be differences in clinical status classification with use of this approach. There are fewer cases above the cutoff for the SP263 assay because the cutoff is higher, at 25% of TCs stained, as compared with 1% for the 28-8 and 22C3 assays. Although the SP142 assay stained fewer TCs and therefore fewer cases were above the threshold on TC score, this was compensated for by six cases rising above the threshold on account of the additional assessment of IC staining.

Although this is an interesting comparison of the assays available, it is very unlikely that in clinical practice, pathologists would choose to assess PD-L1 expression in this way. The ultimate goal is to use the PD-L1 IHC status to inform clinicians and patients of the likelihood of a patient's response to and outcome with a particular programmed cell death 1 or PD-L1 inhibitor. The debate about use of a PD-L1 IHC assay must be informed by relevant clinical trial data for the drug being considered; consequently, the threshold relevant to the drug is the key parameter, not the threshold relevant to the assay. To assign patients above a threshold for treatment, pathologists should not use an alternative cutoff or definition different from that used to define treatment responses in clinical trials for a particular drug.

The comparison of assay performance in terms of assigning a case above or below a cutoff relevant to the proposed drug is therefore more interesting. These comparisons are presented in Table 2. It is important to note that the comparisons here are based on the scores of the expert pathologist trained in each scoring method. Thus, the scores are the best case scenario not confounded by pathologist variation and therefore better reflect variances in the assay itself. The data in Table 2 show that for the 22C3 and SP263 assays, the PD-L1-based classification of cases against all three alternate thresholds agrees in more than 85% of cases when compared with the classification according to the reference assay algorithm. The agreement with the reference assay results is similar (>85%) for the 28-8 assay when classified by either 1% or 25% TPS; however, when the 28-8 assay was assessed according to the SP142 TC1/IC1 algorithm, slightly fewer cases (81.6%) were concordantly classified against the reference SP142 assay. These assays are not identical; it is unrealistic in this scenario to expect 100% concordance, and indeed, a performance of greater than 85% concordance may be clinically acceptable.^{27,28} It is not unexpected that if classification of TC scores according to the 1% and 25% thresholds, as defined respectively by 28-8 or 22C3 and SP263, is attempted by using the SP142 assay, fewer cases (approximately 64%) match the index assay's PD-L1 classification. This reflects the lower levels of TC

staining seen when the SP142 assay is used. The authors acknowledge that the current study does not include an examination of intra-assay variability in the tumor categorization of “above or below” thresholds. Thus, the greater than 85% reproducibility with comparison of three of the four assays might be affected by an intra-assay variation.

It is tempting to view the data shown for the relative similarities between three of the four assays as signaling interchangeability of these assays for clinical use. This study cannot lead to such a conclusion. This feasibility study has examined the technical performance of the PD-L1 IHC assays in a small group of cases that were read by experts in the field at a single test site. Much more data need to be generated, repeating some if not all, of the comparative analyses shown in this article by using real-world small biopsy samples read by a larger number of pathologists, notwithstanding the need for those pathologists to have some training in reading the respective assays. It would be premature to deny that such interchange of assays will remain forever impossible, but currently there are insufficient data to make such a recommendation. It is worth noting that in this small cohort, approximately 15% of patients would *not* have been assigned to a treatment, at least in the context of a companion diagnostic assay, had the alternative similar assay been used. How far short of the technical performance of the trial-validated, definitive standard, drug-associated assay can an alternative testing approach fall but still be accepted in clinical use? It also needs to be remembered that there are no data on the ability of an alternative assay to deliver predictions of treatment response and clinical outcome validated in respective trials.

It is important to note that each PD-L1 IHC assay used in this study was developed in the context of a specific clinical program as a complete system solution including a primary antibody clone, detection reagents, a staining platform, and a software protocol.^{21,29,30} Scoring and interpretation guidelines were developed to identify responding populations for unique drugs and biologic

hypotheses. By virtue of their use in clinical trials, these assays have each been validated by a correlation with patient outcome after being analytically validated as required by the FDA through the premarket application approval process. The stringency of this type of development program is worth bearing in mind. Because these primary antibody clones (as well as others not described here) are available on the market as stand-alone reagents, it is highly likely that laboratory-developed tests (LDTs) will be used as substitutes for the validated assays included in this study. In the absence of rigorous analytical and clinical validation, these LDTs should not be considered equivalent to the companion or complementary diagnostic assays described here for their ability to direct treatment decisions, even if they have been developed using the same antibody clone. LDTs are not developed in concert with a therapeutic development program, and therefore, the cutoff values associated with them are not calibrated to clinical outcomes with use of specific drugs. Also, the staining platforms for the specific assays are defined and specific with the Dako antibodies used with the Dako Link 48 platform and the Ventana antibodies used with the Ventana Benchmark platform. Most recently, a study published by Neuman et al. showed that the 22C3 assay can be successfully used on the Ventana Benchmark platform when appropriate protocol adaption is applied.³¹

In this era of evidence-based medical practice, the exciting development of immunotherapy for lung cancer has been tempered by a complex and hitherto unfamiliar biomarker testing scenario. The only PD-L1 IHC testing approach for which the likely outcome for patients is known is for the test that was performed and validated in the respective clinical trials for each drug (scenario A, Table 3). Any deviation from that practice brings with it unknown consequences for the patient. The case for never deviating from the treatment group-defining cutoff for each drug has been made (data presented in the heat map [Fig. 4] and scenario C, Table 3). It may become acceptable to read some of the

Table 3. Possible Combinations of Drug, PD-L1 Staining Assay, and Slide Scoring Algorithm Use

Scenario	Drug	PD-L1 Assay	Definition of Treatment-Selected Group	Clinical Outcome of Test	Risk to Patient
A	By choice	Assay validated for drug in trial	Definition validated in trial for drug of choice	Predictable on the basis of trial data	Known
B	By choice	Any trial-validated assay	Definition validated in trial for drug of choice	Uncertain	Not known
C	By choice	Any trial-validated assay	Definition validated with the assay	Very uncertain	Not known
D	By choice	LDT using any clone	Unknown	Very uncertain	Not known

PD-L1, programmed death ligand 1; LDT, laboratory-developed test.

assays in a way that would allow alternative cutoffs to be applied to the PD-L1 IHC staining assessment (see Table 2, and scenario B, Table 3), but there are currently no data to validate this practice. At least with this latter approach, the technical outcome of the staining test should be ensured by using a standardized set of reagents and staining platform. Far more uncertain in terms of technical and clinical performance will be the outcomes from the use of nonstandard LDTs (scenario D, Table 3), as has been demonstrated for other IHC biomarkers.^{32,33}

Although the current feasibility study has demonstrated similarities and differences between different PD-L1 IHC assays, it has not addressed specificity and sensitivity of the assays nor clinical outcome comparisons resulting from the classification of a patient's tumor PD-L1 status. The study also did not address how differences in specimen type or preanalytical factors across laboratories (such as fixatives and tissue processing methods) may affect the performance of the assays.³⁴ An ongoing phase 2 of this study includes a much larger set of tumors with comparisons of large specimens versus small biopsy specimens, and even cytologic cell blocks and, potentially, reproducibility of the assay scoring algorithms between pathologists. It is important to emphasize, however, that clinically meaningful comparisons of the PD-L1 assay systems may require therapeutic outcome data, which is a goal that may only be obtainable in the postmarket setting.

In conclusion, the Blueprint Project has revealed that three PD-L1 IHC assays are aligned with regard to PD-L1 expression on TCs, whereas one assay consistently had fewer TCs expressing PD-L1. All the assays demonstrated PD-L1 expression on ICs, but with greater variance than expression on TCs. Substituting cutoffs for the validated cutoff for PD-L1 expression on TCs resulted in changes in the PD-L1 classification for many cases, and a subsequent decrease in overall agreement compared with the validated reference. It is important to note that this is a feasibility study on a small number of cases using expert observers on single assays, and that no results of inferential statistical analyses were obtained owing to the nature of the study's design. A further limitation of this study is that the sample was chosen to represent the range of levels of PD-L1 expression, rather than a representative cohort.

Larger studies using more cases and more pathologists are warranted. On the basis of these limited data, the authors conclude that (1) the thresholds used to select a specific treatment should never be interchanged and (2) more data are required to inform on the validity of using alternative staining assays on which to read different, specific therapy-related PD-L1 cut points.

Acknowledgments

The authors acknowledge the role of the U.S. Food and Drug Administration, American Association for Cancer Research, and American Society of Clinical Oncology in cosponsoring a public workshop titled Complexities in Personalized Medicine: Harmonizing Companion Diagnostics Across a Class of Targeted Therapies, which was held on March 24, 2015, in Washington, D.C., and which provided a public launch pad for the Blueprint Project. In particular, the authors acknowledge Drs. Elizabeth Mansfield, Pamela Bradley, and Reena Philip from U.S. Food and Drug Administration for their thoughtful input, suggestions, and advice in the inception and design of this study. The authors also acknowledge Dr. Jorge Martinalbo of the European Medicines Agency and Chaitali Banerjee at Ventana Medical Systems, Roche Tissue Diagnostics, for their support of this project. The authors thank Dr. Kenichi Suda and Ms. Kristine A. Brovsky at the Division of Medical Oncology, University of Colorado Anschutz Medical Campus, for helping with preparation of the manuscript.

References

1. Gettinger SN, Kowanetz M, Koeppen H, et al. Molecular, immune and histopathological characterization of NSCLC based on PDL1 expression on tumor and immune cells and association with response to the anti-PDL1 antibody MPDL3280A [abstract]. *J Clin Oncol*. 2015;33(suppl):3015.
2. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. 2015;372:2018-2028.
3. Herbst RS, Baas P, Kim DW, 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;387:1540-1550.
4. 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;373:1627-1639.
5. Rizvi NA, Mazieres J, Planchard D, et al. Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol*. 2015;16:257-265.
6. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015;373:123-135.
7. Rizvi NA, Brahmer JR, Ou S-HI, et al. Safety and clinical activity of MEDI4736, an anti-programmed cell death-ligand 1 (PD-L1) antibody, in patients with non-small cell lung cancer (NSCLC) [abstract]. *J Clin Oncol*. 2015;33(suppl):8032.
8. Spigel DR, Chaft JE, Gettinger SN, et al. Clinical activity and safety from a phase II study (FIR) of MPDL3280A (anti-PDL1) in PD-L1-selected patients with non-small

- cell lung cancer (NSCLC) [abstract]. *J Clin Oncol*. 2015;33(suppl):8028.
9. Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multi-centre, open-label, phase 2 randomised controlled trial. *Lancet*. 2016;387:1837-1846.
 10. 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.
 11. Merck and Company, Inc. FDA approves Merck's KEYTRUDA (pembrolizumab) in metastatic NSCLC for first-line treatment of patients whose tumors have high PD-L1 expression (tumor proportion score [TPS] of 50 percent or more) with no EGFR or ALK genomic tumor aberrations. <http://www.mercknewsroom.com/news-release/prescription-medicine-news/fda-approves-mercks-keytruda-pembrolizumab-metastatic-nsclc>. Accessed November 10, 2016.
 12. U.S. Food and Drug Administration. Nivolumab (Opdivo). <http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm436566.htm>. Accessed November 10, 2016.
 13. U.S. Food and Drug Administration. FDA approves Keytruda for advanced non-small cell lung cancer. <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm465444.htm>. Accessed November 10, 2016.
 14. Roche: Media release: Roche receives FDA approval for novel PD-L1 biomarker assay. <http://www.ventana.com/pd-l1-biomarker-assay-news>. Accessed November 10, 2016.
 15. Dako PD-L1 IHC 28-8 pharmDX IFU. <http://www.agilent.com/en-us/products/pharmdx/pd-l1-ihc-28-8-overview>. Accessed November 10, 2016.
 16. Dako PD-L1 IHC Pharm testing. <http://www.agilent.com/en-us/products/pharmdx/pd-l1-ihc-22c3-pharmdx-testing>. Accessed November 10, 2016.
 17. Kerr KM, Tsao MS, Nicholson AG, et al. Programmed death-ligand 1 immunohistochemistry in lung cancer: in what state is this art? *J Thorac Oncol*. 2015;10:985-989.
 18. Cagle PT, Bernicker EH. Challenges to biomarker testing for PD-1/PD-L1 checkpoint inhibitors for lung cancer. *Arch Pathol Lab Med*. 2015;139:1477-1478.
 19. Kerr KM, Hirsch FR. Programmed death ligand-1 immunohistochemistry: friend or foe? *Arch Pathol Lab Med*. 2016;140:326-331.
 20. U.S. Food and Drug Administration. A blueprint proposal for companion diagnostic comparability. <http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM439440.pdf>. Accessed November 10, 2016.
 21. 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;23:541-549.
 22. 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;24:392-397.
 23. 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.
 24. Ilie M, Long-Mira E, Bence C, et al. Comparative study of the PD-L1 status between surgically resected specimens and matched biopsied of NSCLC patients reveal major discordances: a potential issue for anti-PD-L1 therapeutic strategies. *Ann Oncol*. 2016;27:147-153.
 25. Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol*. 2016;29:1165-1172.
 26. Ruschoff J, Kerr KM, Grote HJ, et al. Reproducibility of immunohistochemical scoring for epidermal growth factor receptor expression in non-small cell lung cancer: round robin test. *Arch Pathol Lab Med*. 2013;137:1255-1261.
 27. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*. 2014;138:1432-1443.
 28. Lin F, Chen Z. Standardization of diagnostic immunohistochemistry: literature review and geisinger experience. *Arch Pathol Lab Med*. 2014;138:1564-1577.
 29. Dolled-Filhart M, Roach CM, Toland G, et al. Development of a PD-L1 immunohistochemistry (IHC) assay for use as a companion diagnostic for pembrolizumab (MK-3475) in non-small cell lung cancer (NSCLC) [abstract]. *J Clin Oncol*. 2015;33(suppl):11065.
 30. Rebelatto M, Mistry A, Sabalos C, et al. Development of a PD-L1 companion diagnostic assay for treatment with MEDI4736 in NSCLC and SCCN patients [abstract]. *J Clin Oncol*. 2015;33(suppl):8033.
 31. Neuman T, London M, Kania-Almoq J, et al. A harmonization study for the use of 22C3 PD-L1 immunohistochemical staining on Ventana's platform. *J Thorac Oncol*. 2016;11:1863-1868.
 32. Ibrahim M, Parry S, Wilkinson D, et al. ALK Immunohistochemistry in NSCLC: discordant staining can impact patient treatment regimen. *J Thorac Oncol*. 2016;11:2241-2247.
 33. Wolff AC, Hammond ME, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathology Lab Med*. 2007;131:18-43.
 34. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther*. 2015;14:847-856.