ORIGINAL ARTICLE

# Multicenter Comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) Assays to Test PD-L1 Expression for NSCLC Patients to Be Treated with Immune Checkpoint Inhibitors

Antonio Marchetti, MD, PhD,[a,*] Massimo Barberis, MD,[b] Renato Franco, MD,[c] Graziano De Luca, MD,[a] Maria Vittoria Pace, PhD,[a] Stefania Staibano, MD,[d] Marco Volante, MD,[e] Fiamma Buttitta, MD, PhD,[a] Elena Guerini-Rocco, MD,[b] Luisella Righi, MD,[e] Tommaso D'antuono, BTech,[a] Giorgio V. Scagliotti, MD,[e] Carmine Pinto, MD,[f] Gaetano De Rosa, MD,[d] Mauro Papotti, MD[e]

[a]*Center of Predictive Molecular Medicine, Center for Excellence on Aging and Translational Medicine, University of Chieti-Pescara, Chieti, Italy*
[b]*Division of Pathology, European Institute of Oncology, Milan, Italy*
[c]*Pathology Unit, University of Campania Luigi Vanvitelli, Naples, Italy*
[d]*Pathology Section, Department of Advanced Biomedical Sciences, University Federico II of Naples, Naples, Italy*
[e]*Department of Oncology, University of Turin, Torino, Italy*
[f]*Medical Oncology, Arcispedale S. Maria Nuova Hospital-IRCCS, Reggio Emilia, Italy*

Received 15 March 2017; revised 22 April 2017; accepted 12 July 2017
Available online - 14 August 2017

## ABSTRACT

**Introduction:** Among the several agents targeting the programmed cell death 1 (PD-1) pathway, pembrolizumab is currently the only one approved for the treatment of patients with NSCLC in association with a companion diagnostic assay, the anti–PD-L1 immunohistochemical (IHC) 22C3 PharmDx (Agilent Technologies, Santa Clara, CA) using the Dako Autostainer (Dako, Carpinteria, CA). However, the Dako platform is not present in each pathology department, and this technical limitation is a major problem for the diffusion of the PD-L1 IHC predictive test for pembrolizumab.

**Methods:** The Italian Society of Anatomic Pathology and Cytopathology and the Italian Association of Medical Oncology in an independent, multicenter study compared the in vitro diagnostics PD-L1 IHC 22C3 pharmDx test (Agilent) on the Dako Autostainer and the in vitro diagnostics Ventana PD-L1 (SP263) test on the Ventana BenchMark platform (Ventana Medical Systems, Tucson, AZ). Using serial sections from tissue microarrays, 100 lung adenocarcinomas were locally stained and scored in four centers with the same antibody batches.

**Results:** A high analytical correlation (more than 90% at the lower 95% confidence interval [CI] value) between PD-L1 expression levels obtained with the 22C3 and SP263 assays was observed. At the proposed clinically relevant cutoffs ($\geq$50% and $\geq$1%), the overall concordances between 22C3 and SP263 data were 0.99 (95% CI: 0.96–1) and 0.80 (95% CI: 0.68–0.91), respectively. The lower agreement between data obtained with the 22C3 and SP263 clones at the cutoff of 1% or higher was mainly related to the lower (about 80%) interrater agreement at this cutoff with each clone.

**Conclusions:** These results indicate a high correlation between PD-L1 IHC expression data obtained with the Agilent PD-L1 IHC 22C3 pharmDx and the Ventana PD-L1 (SP263) tests in NSCLC and suggest that the two assays could be utilized interchangeably as an aid to select patients for first-line and second-line treatment with pembrolizumab and potentially with other anti–PD-1/PD-L1 checkpoint inhibitors.



*Keywords:* Non–small cell lung cancer; PD-1; PD-L1; Immunohistochemistry; Immunotherapy

***Corresponding author.**

**Disclosure: The authors declare no conflict of interest.**

**Address for correspondence: Antonio Marchetti, MD, PhD, Center of Predictive Molecular Medicine, Center of Excellence on Aging and Translational Medicine, University of Chieti-Pescara, Via Colle dell'Ara, 66100 Chieti, Italy. E-mail: amarchetti@unich.it**



**ISSN: 1556-0864**

**http://dx.doi.org/10.1016/j.jtho.2017.07.031**

## Introduction

The immune checkpoint inhibitors anti–programmed cell death 1 (PD-1)/programmed cell death ligand 1 (PD-L1) (anti–PD-1/PD-L1) are currently changing the approach to treatment of patients with (NSCLC). Over the last 2 years, the U.S. Food and Drug Administration (FDA) has granted approval to the anti–PD-1 inhibitors, nivolumab (OPDIVO, Bristol-Myers Squibb, New York, New York)[1] and pembrolizumab (KEYTRUDA, Merck Sharp and Dohme, Hoddesdon, United Kingdom)[2] and the anti–PD-L1 inhibitor atezolizumab (TECENTRIQ, Genentech Oncology, South San Francisco, CA)[3] for the treatment of patients with advanced NSCLC with progression during or after first-line therapy. Nivolumab and pembrolizumab have been approved by the European Medicines Agency (EMA) for the same indication.[4,5] Recently, both the European and U.S. agencies have extended the recommendations for pembrolizumab to the first-line therapy of patients with advanced NSCLC.[6,7] In addition, durvalumab (MEDI4736, AstraZeneca, Wilmington, DE) and avelumab (MSB0010718C, Merck KGaA and Pfizer, Darmstadt, Germany) are being investigated for the treatment of NSCLC.[8–10]

The expression of PD-L1 by means of immunohistochemistry represents the most validated predictor of response to anti–PD-1/PD-L1 inhibitors. Although it is still a matter of controversy, results from clinical trials and recent pooled and meta-analyses have shown enhanced clinical benefit from anti–PD-1/PD-L1 therapy in patients with PD-L1–positive tumors.[11–15] Together with anti–PD-1/PD-L1 inhibitors, different immunohistochemical (IHC) assays have been developed to assess the expression of PD-L1, and different antibodies, clones, platforms, score systems, and cutoff values have been introduced for and linked to a specific inhibitor.[16] In particular, PD-L1 IHC 28-8 PharmDx (Agilent Technologies, Santa Clara, CA/Dako Carpinteria, CA) has been labeled a complementary diagnostic of nivolumab on the basis of clinical evidence suggesting that patients affected by PD-L1–positive NSCLC (with $\geq$1% of immunoreactive tumor cells) had a higher clinical benefit from this treatment.[17] The complementary diagnostic Ventana PD-L1 (SP142) assay (Ventana Medical Systems, Tucson, AZ) has been approved for therapy with atezolizumab in considering of the enhanced survival observed in patients with NSCLC with at least 50% of tumor cells expressing PD-L1 or at least 10% of the tumor area occupied by PD-L1–expressing tumor-infiltrating immune cells.[18] Although these assays may provide useful information to clinicians as complementary diagnostics, neither the PD-L1 IHC 28-8 PharmDx nor Ventana PD-L1 (SP142) assay are strictly required for treatment with nivolumab or atezolizumab, respectively. Finally, the Ventana SP263 clone is still under the FDA regulatory process and has been developed for treatment with durvalumab by using a positivity cutoff of 25% or more tumor cells.[12] Currently, only the PD-L1 IHC 22C3 pharmDx assay has been approved by FDA and recommended by the EMA as a companion diagnostic for pembrolizumab. Essentially, pembrolizumab is the only PD-1/PD-L1 inhibitor that has an FDA and EMA indication restricted to PD-L1–positive tumors.[6] The PD-L1 IHC 22C3 pharmDx test was designed to detect PD-L1 in formalin-fixed, paraffin-embedded NSCLC samples by using mouse monoclonal anti–PD-L1 clone (22C3) and the EnVision FLEX visualization system on Autostainer Link 48 (Agilent).[19] However, not every pathology laboratory has the Dako Autostainer. This represents a major drawback for diffusion of the PD-L1 IHC test in the clinical practice. A harmonization study to analyze the results obtained with the 22C3 clone on the Dako Autostainer and Ventana BenchMark platforms has been recently presented.[20] The overall concordance reported in this study was 85% to 87% and was limited to strongly positive tumors. Indeed, in all investigated cases with weak positivity (tumor proportion score of 1%–49%) the results were discordant. An additional limitation of this approach is that the test was used under off-label (not in vitro diagnostics [IVD]) conditions.

The IVD PD-L1 tests, Ventana PD-L1 (SP142) and the Ventana PD-L1 (SP263) assays have been developed for use on the Ventana BenchMark platforms, which are also largely present in pathology laboratories. Previous multicenter studies for the harmonization of clinically tested diagnostic antibodies have clearly shown that the immunostaining of NSCLC samples with the SP142 clone is significantly weaker and the percentage of positive tumor cells is lower than that obtained with the 22C3 and SP263 clones.[21–24]

To investigate the possibility of using a PD-L1 test alternative to the PD-L1 IHC 22C3 pharmDx assay for those laboratories that do not have the Dako platform, we designed a multicenter study to assess the technical procedures and interpretation ability in four different pathology units with specific expertise in both thoracic pathology and immunohistochemistry. The study aimed to compare the PD-L1 expression evaluated by the IHC 22C3 PharmDx on the Dako Autostainer platform with that observed by the SP263 on the Ventana BenchMark platform in a large number of lung adenocarcinomas (ADC).

## Materials and Methods

### *Organization of the Multicenter Study* Anti-*PD-L1 IHC 22C3*

The Italian Society of Anatomic Pathology and Cytopathology and the Italian Association of Medical Oncology identified a board of Italian pathologists and

oncologists who were appointed to design the project and are coauthors of this study. Four surgical pathology departments (Center of Predictive Molecular Medicine, Center of Predictive Molecular Medicine, Center for Excellence on Aging and Translational Medicine, University of Chieti, Italy; Division of Pathology and Laboratory Medicine, European Institute of Oncology, Milan, Italy; Pathology Division, Department of Oncology, University of Turin, Turin, Italy; and Pathology Division, Departments of Pathology, University of Naples Federico II and University of Campania Luigi Vanvitelli, Naples, Italy) were identified as referral centers for the PD-L1 multicenter study. All the selected centers had both DAKO Autostainer and Ventana BanchMark IHC platforms. The centers have been anonymized as center A, center B, center C, and center D.

### *Tissue Samples and TMA Preparation*

A total of 100 formalin-fixed, paraffin embedded samples from consecutive resected lung adenocarcinomas collected during the period 2014–2016 were used for the preparation of tissue microarrays (TMAs). Lung tumors specimens were provided from three of the four centers participating in the study (centers A, B, and C). In each case, the two representative areas with the highest percentage of vital neoplastic cells were selected on 4-$\mu$m-thick, hematoxylin and eosin–stained sections and labeled with a permanent marker. Corresponding areas on paraffin blocks were numbered, captured with a 2-mm puncher on a semi-automatic tissue micro arrayer (Galileo TMA CK3500B, Integrated Systems Engineering Srl, Milan, Italy), and transferred to recipient TMA blocks arranged in pairs corresponding to single cases to allow an easy comparison. A core of human placental tissue was also placed in each recipient block to provide a positive control and a reference for orienting the slides under the microscope. From eight TMA blocks, for a total of 200 cores, 20 (4-$\mu$m-thick) sections per block were obtained and sent to the four referral centers for subsequent IHC staining and analysis. Each center alternatively received contiguous sections to minimize differences in tumor morphology. Unstained sections were prelabeled with a letter indicating the corresponding block (A, B, C, D, E, F, G, and H), as were the clones to be used for the IHC staining.

### *IHC Analysis*

In each center, IHC analysis was conducted with IVD versions of the PD-L1 IHC 22C3 pharmDx and the Ventana PD-L1 (SP263) assays on the DAKO Autostainer Link 48 and Ventana BanchMark platforms, respectively. The detection and quantification of the percentage of immunoreactive tumor cells was performed according to the manufactures recommendations. Briefly, neoplastic cells were considered positive when any cell membrane staining (partial or complete) was present, ignoring pure cytoplasmic immunoreaction. Staining on immune cells was also disregarded. Quantification of immunoreactive neoplastic cells was obtained by evaluating the ratio between stained carcinoma cells and all viable carcinoma cells. The IHC staining of neoplastic cell was assessed in each (2-mm) TMA core of the same case, and the mean percentage of PD-L1 expression between the two cores was reported. To maximize the standardization of the scoring procedures, a miniguide summarizing the Agilent and Ventana evaluation protocols was sent to the four centers before the analytical process. The senior pathologist in each center evaluated the IHC staining. For three centers (A, B, and C), the pathologist had attended international training workshops on the detection of PD-L1 immunoreaction as implemented by Agilent and Ventana; from now on, we will refer to them as *trained centers.* The referring pathologist of the fourth center (D) had not attended any PD-L1–specific training workshop at the time of this study.

### *Statistical Analysis*

PD-L1 expression values measured in the four pathology centers for each of the two IHC assays were investigated for correlation by using the intraclass correlation coefficient.[25] The agreement among raters at the clinically relevant PD-L1 expression cutoff levels for pembrolizumab ($\geq$1% and $\geq$50%) were assessed by the weighted kappa ($\kappa$) of Fleiss's $\kappa$ interrater agreement analysis, as appropriate.[26] The relationship between PD-L1 expression levels evaluated by the 22C3 and SP263 assays was calculated using the concordance correlation coefficient, Pearson's precision analysis.[27] At the clinically relevant cutoff, the correlation between data generated in the different centers by the 22C3 and SP263 clones was calculated using the weighed $\kappa$ or Light's $\kappa$ statistics, as appropriate.[26] A correlation higher than 90% is a typical within-assay agreement for IHC.[28] A $p$ value less than 0.05 was considered significant. All statistical analyses were performed using IBM SPSS statistics software, version 22 (IBM Corp., Armonk, NY).

## Results

Scoring of the 22C3 and SP263 IHC assays showed qualitative differences in the staining patterns. Staining intensity was slightly stronger and more dense (less granular) with the SP263 clone. Although intensity is not included in the PD-L1 IHC scoring system, a higher intensity can make it easier to recognize the staining at low magnification. Therefore, the percentage of positive cells at low magnification usually seemed to be higher with the SP263 clone. However, at a higher magnification,

when cells having weak and incomplete membrane staining were also scored positive, as otherwise recommended by the manufacturers of both assays, a very similar percentage of positive cells was obtained in each case with the two IHC tests. A paradigmatic example of IHC staining with the 22C3 and SP263 clones is shown in Figure 1.

The distribution of PD-L1 expression levels detected by the 22C3 and SP263 tests in each of the four participating pathology centers are reported in the dot-and-line diagram of Figure 2. The intraclass correlation coefficients for the 22C3 and SP263 clones were 0.973 (95% CI [confidence interval]: 0.964–0.981) and 0.968 (95% CI: 0.957–0.977), respectively. When only the three trained centers were considered, the intraclass correlation coefficients for data obtained with the 22C3 and SP263 clones were 0.976 (95% CI: 0.966–0.983) and 0.977 (95% CI: 0.969–0.984), respectively.

At a cutoff of 50% or higher, the Fleiss's $\kappa$ interrater agreement values were 0.931 (95% CI: 0.851–1.011) and 0.942 (95% CI: 0.862–1.022) for data obtained with the 22C3 and SP263 clones, respectively. Restricted to the three trained centers, the agreement values were 0.946 (95% CI: 0.833–1.059) and 0.973 (95% CI: 0.860–1.087) for the 22C3 and SP263 data, respectively (Table 1).

At the cutoff of 1% or higher, the Fleiss's $\kappa$ interrater agreement values were 0.754 (95% CI: 0.674– 0.834) and 0.798 (95% CI: 0.718–0.878) for the 22C3 and SP263 data, respectively. Among the three trained centers, the Fleiss's $\kappa$ interrater agreement values were 0.769 (95% CI: 0.655–0.882) and 0.82 (95% CI: 0.709–0.936) for the 22C3 and SP263 data, respectively.

The correlation between the 22C3 and SP263 expression levels in each of the four centers is shown in scatter diagrams (Fig. 3*A*–*D*). Each point in the diagram

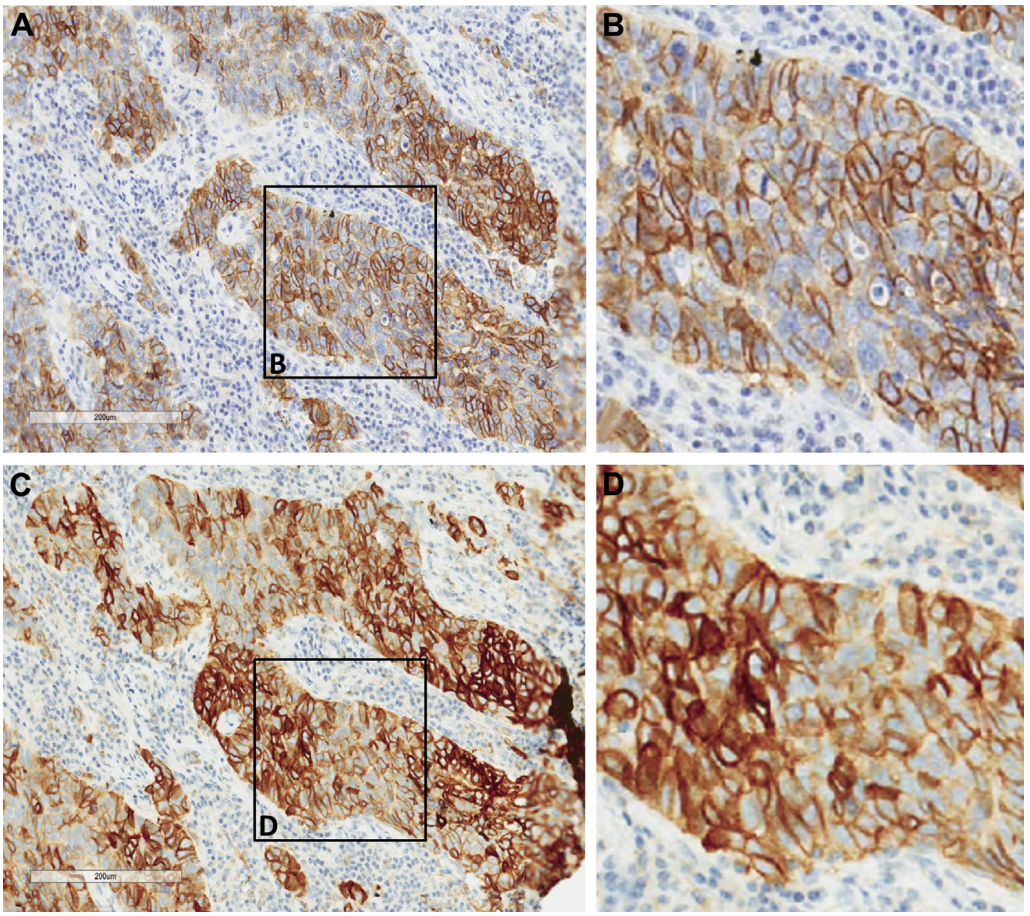


**Figure 1.** A typical example of immunohistochemical staining with the programmed death ligand 1 immunohistochemical 22C3 pharmDx test (Agilent) on the Dako Autostainer (*A*) and the in vitro diagnostics Ventana programmed death ligand 1 (SP263) test on the on the Ventana BenchMark platform (*C*) is reported. Note that at a low magnification (×100) (*A* and *C*), the percentage of positive tumor cells with the SP263 clone seems to be higher owing to the slightly higher intensity of the staining. However, at higher magnification (×200-400), if a weak and incomplete cell membrane staining is considered according to the manufacturer recommendations, the two tests show a very similar percentage of immunoreactive tumor cells (magnified frames *B* and *D*).

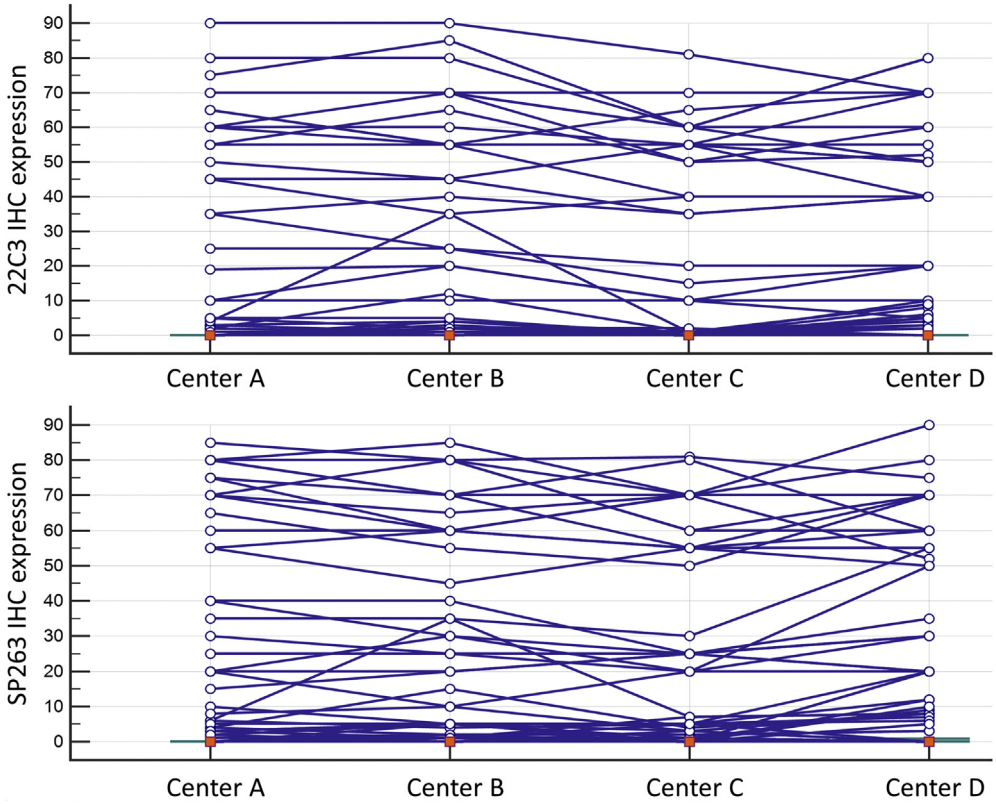


**Figure 2.** Dot-and-line diagram of the distribution and correlation of programmed death ligand 1 (PD-L1) expression levels detected by the PD-L1 immunohistochemistry (IHC) 22C3 pharmDx test (Agilent), on the Dako Autostainer (*top*) and the in vitro diagnostics Ventana PD-L1 (SP263) test on the Ventana BenchMark platform (*bottom*) in each of the four referral pathology centers. Data are often superimposed, mainly for cases with low PD-L1 expression levels.

indicates the percentage of positive tumor cells with the anti–PD-L1 22C3 (on the *x* axis) and SP263 (on the *y* axis) assays for each of the 100 cases evaluated. The concordance correlation coefficients between the expression levels of 22C3 and SP263 were 0.97 (0.95–0.98), 0.97 (0.96–0.98), 096 (0.94–0.97), and 0.89 (0.84–0.92) for centers A, B, C, and D, respectively, as reported in Table 2 (boldface entries). The table also indicates intercenter agreements for all the coder pairs.

A line corresponding to the 50% cutoff for both clones is indicated in Figure 3. Analytical results of all cases scored 50% or higher in at least one center are reported in Table 3. In panel A, cases with a mean 22C3 or SP263 PD-L1 expression of 50% or higher are reported. Note that all cases in panel A were scored positive in at least two of the trained centers. Panel B shows an additional two cases with a mean of PD-L1 expression less than 50%. At the cutoff of 50% or higher, 15 cases (15%) were found to be positive with both clones by center A. Fourteen of these 15 cases (93%) were judged positive by center B and C. However, the discordant case in the last two centers was not the same and had an expression value slightly different from that reported from the other centers (see Table 3, boldface entries). The discordant case for center B (case 3A) was scored 45% with both clones. Similarly, the discordant case for center C (case 2E) was scored 40% with clone 22C3 and 55% with clone SP263. The concordance of data obtained by the pathologist of center D, who was not trained for PD-L1 assessment in

**Table 1.** Interrater Agreement for PD-L1 Expression Data at Different Cutoff Levels

| PD-L1 Expression Cutoff | 22C3 Test (95% CI)[a] | SP263 Test (95% CI)[a] | 22C3 Test vs. SP263 Test[b] |
|---|---|---|---|
| ≥50% | 0.95 (0.83-106) | 0.97 (0.86-1.09) | 0.99 (0.96-1) |
| ≥1% | 0.77 (0.65-088) | 0.82 (0.71-0.94) | 0.80 (0.68-0.91) |

[a]Fleiss's $\kappa$ interrater agreement.
[b]Light's $\kappa$ for all coder pairs.
PD-L1, programmed death ligand 1; CI, confidence interval.

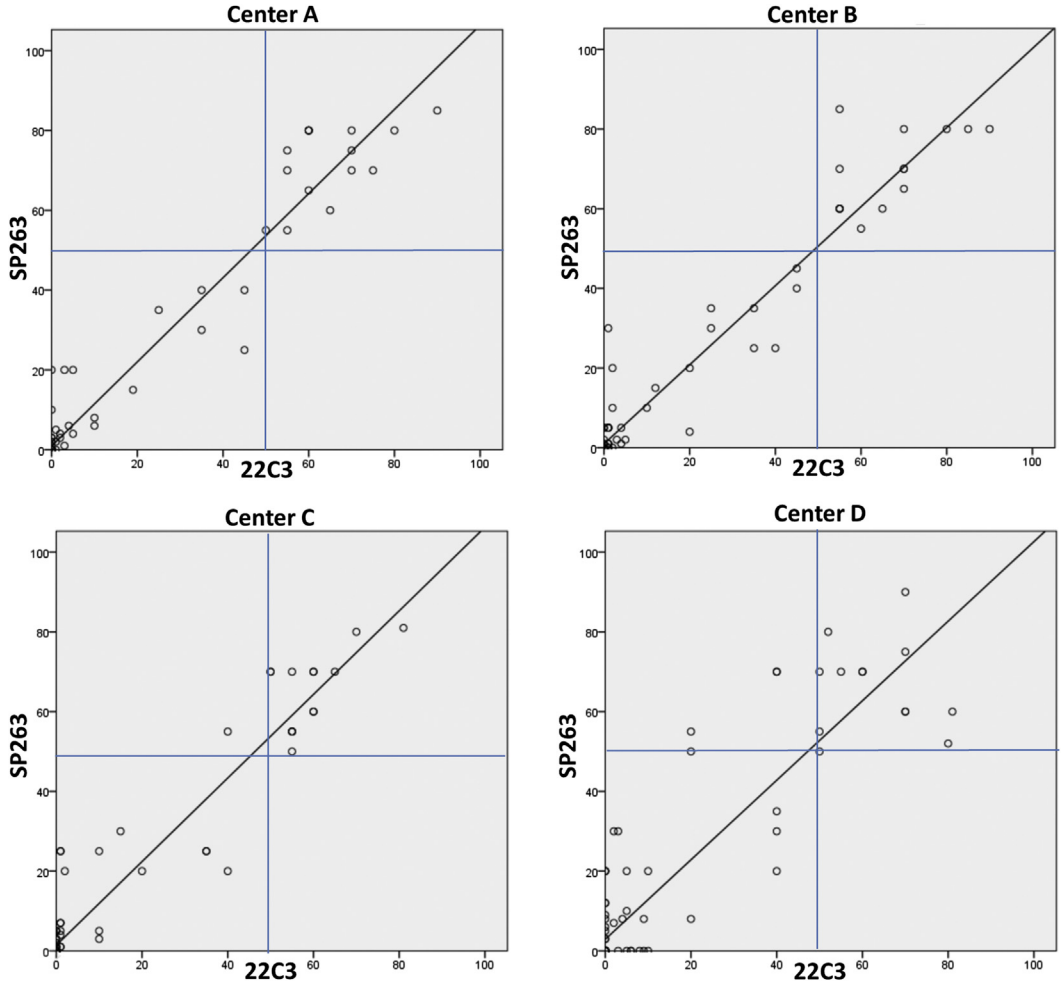


**Figure 3.** Scatter diagrams illustrating the correlation between the programmed death ligand 1 immunohistochemical 22C3 pharmDx test (22C3) and in vitro diagnostics Ventana programmed death ligand 1 (SP263) expression levels in each of the four centers. See the text for details.

international training workshops, was lower: two of the 15 cases in panel A were classified as negative with clone 22C3 (cases 2E and 3H, both scored 40%); in addition, the two cases in panel B that were scored negative with the SP263 clone by the other three centers and with a mean PD-L1 expression level less than 50% were scored positive (cases 6D and 1G, which were scored 50% and 55%, respectively). Overall, center D reported four discordant cases.

At the cutoff of 50% or higher, the concordances (weighted $\kappa$ values) between 22C3 and SP263 data in each center were 1, 1, and 0.96 (95% CI: 0.88–1) for centers A, B, and C, respectively, and 0.844 (95% CI: 0.695–0,992) for center D. The average concordance

**Table 2.** Concordance between PD-L1 Expression Levels Obtained with the 22C3 and SP263 Tests in the Four Pathology Centers

| Center | Center 1 SP263 (95% CI)[a] | Center 2 SP263 (95% CI)[a] | Center 3 SP263 (95% CI)[a] | Center 4 SP263 (95% CI)[a] |
|---|---|---|---|---|
| Center 1 22C3 | **0.97 (0.95-0.98)** | 0.96 (0.95-0.97) | 0.96 (0.94-0.97) | 0.93 (0.90-0.95) |
| Center 2 22C3 | 0.96 (0.94-0.97) | **0.97 (0.96-0.98)** | 0.95 (0.93-0.97) | 0.92 (0.88-0.94) |
| Center 3 22C3 | 0.94 (0.92-0.96) | 0.94 (0.91-0.95) | **0.96 (0.94-0.97)** | 0.90 (0.86-0.93) |
| Center 4 22C3 | 0.94 (0.92-0.95) | 0.94 (0.92-0.96) | 0.95 (0.92-0.96) | **0.89 (0.84-0.92)** |

*Note:* Boldface indicates intracenter concordance.
[a]Concordant correlation coefficient.
PD-L1, programmed death ligand; CI, confidence interval.

**Table 3.** PD-L1 Expression Data Reported from the Four Centers in All Cases with a Score of at Least 50% in at Least One Center

| Panel | Case | Clone | Center A | Center B | Center C | Center D | Mean |
|---|---|---|---|---|---|---|---|
| A | 3A | 22-C3 | 50 | **45** | 55 | 50 | 50 |
| | | SP-263 | 55 | **45** | 55 | 55 | 52.5 |
| | 7A | 22-C3 | 90 | 90 | 80 | 70 | 82.5 |
| | | SP-263 | 85 | 80 | 80 | 75 | 80 |
| | 9A | 22-C3 | 60 | 55 | 65 | 70 | 62.5 |
| | | SP-263 | 80 | 85 | 70 | 90 | 81.25 |
| | 4B | 22-C3 | 70 | 70 | 60 | 60 | 65 |
| | | SP-263 | 70 | 65 | 70 | 70 | 68.75 |
| | 9B | 22-C3 | 55 | 65 | 50 | 52 | 55.5 |
| | | SP-263 | 70 | 60 | 70 | 80 | 70 |
| | 2C | 22-C3 | 70 | 70 | 70 | 70 | 70 |
| | | SP-263 | 80 | 70 | 80 | 60 | 72.5 |
| | 3C | 22-C3 | 55 | 55 | 55 | 55 | 55 |
| | | SP-263 | 75 | 60 | 70 | 70 | 68.75 |
| | 4C | 22-C3 | 55 | 55 | 55 | 50 | 53.75 |
| | | SP-263 | 55 | 60 | 55 | 50 | 55 |
| | 12C | 22-C3 | 60 | 70 | 60 | 50 | 60 |
| | | SP-263 | 80 | 80 | 60 | 70 | 72.5 |
| | 7D | 22-C3 | 70 | 70 | 50 | 60 | 62.5 |
| | | SP-263 | 75 | 70 | 70 | 70 | 71.25 |
| | 2E | 22-C3 | 65 | 55 | **40** | **40** | 50 |
| | | SP-263 | 80 | 70 | 55 | 70 | 68.75 |
| | 7E | 22-C3 | 75 | 85 | 60 | 80 | 75 |
| | | SP-263 | 70 | 80 | 60 | 60 | 67.5 |
| | 12E | 22-C3 | 65 | 55 | 55 | 70 | 61.25 |
| | | SP-263 | 60 | 60 | 55 | 60 | 58.75 |
| | 3F | 22-C3 | 80 | 80 | 60 | 80 | 75 |
| | | SP-263 | 80 | 80 | 70 | 52 | 70.5 |
| | 3H | 22-C3 | 60 | 60 | 55 | **40** | 53.75 |
| | | SP-263 | 65 | 55 | 50 | 70 | 60 |
| B | 6D | 22-C3 | 25 | 25 | 15 | 20 | 21.25 |
| | | SP-263 | 35 | 35 | 30 | **55** | 38.75 |
| | 1G | 22-C3 | 35 | 25 | 20 | 20 | 25 |
| | | SP-263 | 40 | 30 | 20 | **50** | 35 |

*Note:* Details in the test. Boldface indicate discordant values.

(Light's $\kappa$ values) for the three trained centers at the cutoff of 50% or higher was 0.99 (95% CI: 0.96–1) (see Table 1). At the cutoff of 1% or higher, the concordances between 22C3 and SP263 data were 0.83 (95% CI: 0.75–0.88), 0.8 (95% CI: 0.68–0.92), 0.77 (95% CI: 0.64–0.89), and 0.62 (95% CI: 0.46–0.77) for centers A, B, C, and D, respectively. The average agreement for the three trained centers at the cutoff of 1% or higher was 0.80 (95% CI: 0.68–0.91) (see Table 1).

The mean frequencies of lung adenocarcinomas found to be positive with the 22C3 and SP263 tests were 37.3% and 41.3%, respectively, at the cutoff of 1% or higher and 14.3% and 14.7%, respectively, at the cutoff of 50% or higher.

## Discussion

Identification and accurate evaluation of biomarkers are critical steps for targeted therapy of patients with NSCLC, including immunotherapy with anti–PD-1/PD-L1 drugs.[29] A number of clinical trials have recently shown that IHC expression of PD-L1 in tumor tissue correlates with better response rate and clinical outcome after anti–PD-1 treatment with pembrolizumab. In the pivotal KEYNOTE 01/010/024 studies, which led to FDA and EMA approval of pembrolizumab for first- and second-line treatment of patients with NSCLC, the PD-L1 22C3 IHC assay on the Dako Autostainer platform has been used for patient selection.[30–32] Whereas in the FDA approval the drug's administration is linked to an FDA-approved test, the EMA more generally recommends the use of a validated test.[30–32] Because the Dako system is not present in all pathology laboratories, a technical issue suddenly emerged for PD-L1 IHC–based selection of patients to be treated with pembrolizumab. A number of harmonization studies have been conducted to deliver analytical data for the main commercially available

PD-L1 diagnostic assays. Preliminary results have shown promise that at least some of the PD-L1 IHC assays might be complementary.[21–24] However, some considerations that emerged from these previous projects must be taken into account in evaluation of the data: (1) the number of cases examined is still too limited to draw definitive conclusions; (2) most projects were developed to compare sections stained in a central laboratory, so the impact of the preanalytical phase performed in different centers was not assessable; (3) the studies were usually performed on sections of resected tumors, where an accurate comparison of the percentage of immune-reactive neoplastic cells may be more challenging than that attainable on accurately selected tumor areas; (4) several studies have been conducted with research use–only reagents; and (5) some of the projects were conducted in collaboration with commercial companies.

This independent, multicenter study was specifically conceived to compare the IVD Ventana PD-L1 (SP263) test on the Ventana BenchMark platform with the IVD PD-L1 IHC 22C3 pharmDx test on the Dako Autostainer platform on selected tumor areas from a large series of lung adenocarcinomas.

Our results indicate a high analytical correlation—higher than the typical 90% within-assay agreement for IHC—between the expression data obtained by the different observers with each of the two commercially available PD-L1 assays.[28] Indeed, the concordance correlation coefficient among the four raters was 97% with both IHC tests (98% with both tests when the untrained center was excluded).

At the cutoff of 50% or higher, which is used for the selection of patients for first-line treatment with pembrolizumab,[32] the interrater agreement was far higher than 90%, especially when only the three trained centers were considered (95% and 97% for the 22C3 and SP263 tests, respectively).

A different scenario emerged when the same cases were classified with a cutoff of 1% or higher. The interrater agreement decreased to 75% and 80% for the 22C3 and SP263 data, respectively. When only the three trained centers were considered, the agreement was slightly improved (77% and 82% for the 22C3 and SP263 data, respectively). These results clearly indicate that the overall agreement at a cutoff of 1% or higher is lower, near 80%, and can potentially be increased by a specific training. The lower interobserver agreement at the cutoff of 1% that emerged in this study is in keeping with data reported in a recent PD-L1 expression harmonization project that included the 22C3 and SP263 clones.[23] Regardless of the results of currently available clinical studies, these observations should prompt further consideration of the opportunity to adopting such a low cutoff ($\geq$1%) for clinical selection of patients to be treated with anti–PD-L1 drugs. However, in the harmonization study by Scheel et al., the interobserver agreement was similar at all cutoffs.[22]

The present study showed a high analytical correlation (more than 90% at the lower 95% CI value) (see Table 2) between PD-L1 expression levels evaluated by the 22C3 and SP263 tests in the three trained centers. Similarly, at the cutoff of 50% or higher, a high degree of concordance between 22C3 and SP263 data (96%–100%) was observed for the three trained centers. The concordance for center D was only 84%, which is slightly lower than that requested ($\geq$90%) by the College of American Pathologists guidelines for IHC assays.[28] The lower performance of center D was due to the fact that at the cutoff of 50% or higher, discordant results occurred in four cases (see Table 3). Because these analytical data could reflect differences in the preanalytical phase, high-resolution scanned images of all discrepant cases stained with both antibodies in center A were sent to center D. The four cases gave discordant data in center D also when evaluated on digitalized images, clearly indicating that it was a matter of staining interpretation.

At the cutoff of 1% or higher, the concordance between 22C3 and SP263 data in the trained centers was about 80% (77%, 80%, and 83%), whereas in the untrained center the concordance was much lower (62%). These data clearly show the importance of a specific training for pathologists to ensure reproducibility of results across laboratories, especially at a low cutoff level. The lower agreement between the data obtained with the 22C3 and SP263 clones at the cutoff of 1% or higher is mainly attributable to the lower (about 80%) interrater agreement obtained at this cutoff level with each clone, as previously discussed.

An additional finding that emerged from the current study is the prevalence of PD-L1 expression at different cutoffs in a series of 100 consecutive resected lung adenocarcinoma collected from three Italian centers. About 40% of the tumors were PD-L1–positive with both clones at the cutoff of 1% or higher, whereas only about 15% of cases were positive at the cutoff of 50% or higher. These results are in keeping with recently published data on resected lung adenocarcinomas.[33]

In conclusion, our results demonstrate a high correlation between PD-L1 IHC expression data obtained with the 22C3 and SP263 tests in NSCLC. The agreement is excellent at a cutoff of 50% or higher and still within the range of the intraclone agreement for both antibodies at a cutoff of 1% or higher. Although clinical validation for this alternative staining approach is lacking, our data suggest that the PD-L1 IHC 22C3

pharmDx and the VENTANA PD-L1 (SP263) assays could be used interchangeably as an aid to select patients for first-line and second-line treatment with specific targeted agents.

## Acknowledgments

Support for this work was provided by the Italian Ministry of Education, Universities and Research.

## References

1. US Food and Drug and Administration. FDA approved drug products. http://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=125527. Accessed March 2017.
2. US Food and Drug Administration. FDA approves Keytruda for advanced non-small cell lung cancer. https://www.fda.gov/newsevents/newsroom/pressansupnouncements/ucm465444.htm. Accessed October 2, 2015.
3. US Food and Drug Administration. Atezolizumab (TECENTRIQ). https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm525780. Accessed October 18, 2016.
4. European medicines agency. Opdivo. http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion/human/003985/WC500194111.pdf. Accessed September 24, 2015.
5. European medicines agency. Keytruda. http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion/human/003820/WC500209310.pdf. Accessed June 23, 2016.
6. US Food and Drug Administration. Pembrolizumab (KEYTRUDA) checkpoint inhibitor. https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm526430. Accessed October 24, 2016.
7. European medicines agency. Keytruda. http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion/human/003820/WC500218016.pdf. Accessed December 15, 2016.
8. Antonia S, Goldberg SB, Balmanoukian A, et al. Safety and antitumour activity of durvalumab plus tremelimumab in non-small cell lung cancer: a multicentre, phase 1b study. *Lancet Oncol*. 2016;17:299–308.
9. Planchard D, Yokoi T, McCleod MJ, et al. A phase III study of durvalumab (MEDI4736) with or without tremelimumab for previously treated patients with advanced NSCLC: rationale and protocol design of the Arctic study. *Clin Lung Cancer*. 2016;17:232–236.e231.
10. Jerusalem G, Chen F, Spigel D, et al. Oa03.03 javelin solid tumor: OA03.03 JAVELIN solid tumor: safety and clinical activity of avelumab (anti-PD-L1) as first-line treatment in patients with advanced NSCLC. *J Thorac Oncol*. 2017;12:S252.
11. Grigg C, Rizvi NA. PD-L1 biomarker testing for non-small cell lung cancer: truth or fiction? *J Immunother Cancer*. 2016;4:48.
12. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther*. 2015;14:847–856.
13. Abdel-Rahman O. Correlation between PD-L1 expression and outcome of NSCLC patients treated with anti-PD-1/PD-L1 agents: a meta-analysis. *Crit Rev Oncol Hematol*. 2016;101:75–85.
14. Carbognin L, Pilotto S, Milella M, et al. Differential activity of nivolumab, pembrolizumab and mpdl3280a according to the tumor expression of programmed death-ligand-1 (PD-L1): sensitivity analysis of trials in melanoma, lung and genitourinary cancers. *PLoS One*. 2015;10:e0130142.
15. Sacher AG, Gandhi L. Biomarkers for the clinical use of PD-1/PD-L1 inhibitors in non-small-cell lung cancer: a review. *JAMA Oncol*. 2016;2:1217–1222.
16. Gridelli C, Ascierto PA, Barberis MC, et al. Immunotherapy of non-small cell lung cancer: report from an international experts panel meeting of the Italian Association of Thoracic Oncology. *Expert Opin Biol Ther*. 2016;16:1479–1489.
17. US Food and Drug Administration. Premarket approval (PMA). http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P150027. Accessed January 23, 2016.
18. Roche. Ventana. http://www.accessdata.fda.gov/cdrh_docs/pdf16/p160006c.Pdf. Accessed October 21, 2016.
19. Dako. PD-L1 IHC 22C3 pharmDx. https://www.agilent.com/cs/library/packageinsert/public/P03951%20rev%2004.pdf. Accessed October 2, 2016.
20. Neuman T, London M, Kania-Almog J, et al. A harmonization study for the use of 22C3 PD-L1 immunohistochemical staining on Ventana's platform. *J Thorac Oncol*. 2016;11:1863–1868.
21. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. *J Thorac Oncol*. 2017;12:208–222.
22. Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol*. 2016;29:1165–1172.
23. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between programmed cell death ligand-1 diagnostic assays across multiple protein expression cut-offs in non-small cell lung cancer. *Clin Cancer Res*. 2017;23:3585–3591.
24. Adam J, Rouquette I, Damotte D, et al. Multicentric French harmonization study for PD-L1 IHC testing in NSCLC abstract PLO4a. *J Thorac Oncol*. 2016;12:S11–S12.
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
26. Hallgren KA. Computing Inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8:23–34.
27. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255–268.
28. Fitzgibbons PL1, Bradley LA, Fatheree LA, et al. College of American Pathologists Pathology and Laboratory Quality Center. 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.
29. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*. 2012;366:2455–2465.
30. 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.
31. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375:1823–1833.
32. 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.
33. Wu S, Shi X, Sun J, et al. The significance of programmed cell death ligand 1 expression in resected lung adenocarcinoma. *Oncotarget*. 2017;8:16421–16429.