Clinical Utility of the Combined Positive Score for Programmed Death Ligand-1 Expression and the Approval of Pembrolizumab for Treatment of Gastric Cancer

Karina Kulangara, PhD; Nancy Zhang, MD; Ellie Corigliano, PhD; Lindsay Guerrero, MS; Stephanie Waldroup, BSc; Dipeshkumar Jaiswal, MS; Malinka Jansson, MS; Supriya Shah, PhD; Debra Hanks, MD; Jiangdian Wang, PhD; Jared Lunceford, PhD; Mary J. Savage, PhD; Jonathan Juco, MD; Kenneth Emancipator, MD

• ***Context.***—**Regulatory approval of pembrolizumab for treatment of gastric and gastroesophageal junction (G/GEJ) adenocarcinoma required a reproducible scoring method for use of programmed death ligand-1 (PD-L1) protein expression as a companion diagnostic to identify likely responders to therapy.**

Objective.—**To develop an immunohistochemical scoring algorithm that includes PD-L1 expression for tumor and immune cells, that is, the combined positive score.**

Design.—**Four previously treated tumor types in the KEYNOTE-012 and KEYNOTE-028 studies were analyzed descriptively with a version of the PD-L1 immunohistochemical 22C3 pharmDx assay labeled for investigational use only to determine the relative importance of PD-L1 expression in tumor versus immune cells as a biomarker for pembrolizumab response. A combined positive score was developed as a novel scoring method and was compared with the tumor proportion score in cohort 1 from the KEYNOTE-059 study (G/GEJ cancer). External reproducibility was assessed.**

Results.—**Per combined positive score cutoff of 1 or more, the prevalence of PD-L1 expression in patients with G/GEJ cancer was 57.6% (148 of 257 patients), with reasonable enrichment of responses (odds ratio, 2.8). Per tumor proportion score cutoff of 1% or more, prevalence was 12.5% (32 of 257 patients), with minimal enrichment (odds ratio, 1.4). External reproducibility assessments demonstrated interpathologist overall agreement of 96.6% (591 of 612; 95% CI, 94.0%–98.7%) and intrapathologist overall agreement of 97.2% (595 of 612; 95% CI, 95.3%–98.9%).**

Conclusions.—**Combined positive score is a robust, reproducible PD-L1 scoring method that predicts response to pembrolizumab in patients with G/GEJ cancer. This novel scoring method supported US Food and Drug Administration approval of pembrolizumab as third-line therapy for G/GEJ cancer and has facilitated investigation in other indications.**

(*Arch Pathol Lab Med.* 2019;143:330–337; doi: 10.5858/arpa.2018-0043-OA)

Programmed death ligand-1 (PD-L1) protein expression in tumor cells has proved to be an invaluable biomarker in non–small cell lung cancer (NSCLC). PD-L1 expression as measured by the tumor proportion score (TPS) supported the US Food and Drug Administration (FDA) granting of breakthrough status to pembrolizumab and to the subsequent accelerated approval of pembrolizumab for the treatment of NSCLC, including approval of PD-L1 immunohistochemical (IHC) 22C3 pharmDx (Dako North America, Agilent Technologies, Carpinteria, California) as the companion diagnostic. More recently, use of this companion diagnostic enabled approval of pembrolizumab monotherapy as first-line therapy for NSCLC, with pembrolizumab remaining the only immunotherapy for this indication.[1]

Although the TPS method has been shown to identify patients who have PD-L1$^+$ NSCLC tumors and who are likely to respond to anti–PD-L1/anti–programmed death receptor-1 (PD-1) therapy,[2–6] measurement of tumor PD-L1 expression alone might not be as suitable a predictive biomarker for many other indications.[7] The potential for immune cell PD-L1 expression to have an important role in the prediction of response to checkpoint inhibitors has been a key area of interest. The modified proportion score, which was used for the 22C3 PD-L1 prototype assay,[8] is based on a mixture of tumor and immune cell expression. That assay was used to enroll patients in the KEYNOTE-012

Accepted for publication April 23, 2018.
Published online July 20, 2018.

From the Division of Companion Diagnostics, Dako North America, Agilent Technologies, Carpinteria, California (Drs Kulangara, Zhang, Corigliano, Shah, and Hanks; Mss Guerrero, Waldroup, and Jansson; and Mr Jaiswal); and Merck Research Laboratory, Merck & Co, Inc, Kenilworth, New Jersey (Drs Wang, Lunceford, Savage, Juco, and Emancipator). Dr Corigliano is now with Merck & Co, Inc.

Drs Zhang, Corigliano, and Shah and Mss Guerrero and Jansson are employees of Agilent Technologies. Drs Kulangara and Hanks, Ms Waldroup, and Mr Jaiswal are employees of Agilent Technologies and own stock and hold stock options in the company. Ms Waldroup has pending patents broadly relevant to the work described. Dr Wang is an employee of Merck & Co, Inc. Drs Lunceford, Savage, Juco, and Emancipator are employees of Merck & Co, Inc, and own stock with this company.

Funding for this study was provided by Merck & Co., Inc. (Kenilworth, New Jersey).

Corresponding author: Kenneth Emancipator, MD, Merck Research Laboratory, Merck & Co, Inc, 2000 Galloping Hill Rd, Kenilworth, NJ 07033 (email: kenneth.emancipator@merck.com).

(ClinicalTrials.gov, NCT01848834) and KEYNOTE-028 (ClinicalTrials.gov, NCT02054806) clinical trials, as well as in expansion cohorts in KEYNOTE-001 (ClinicalTrials.gov, NCT01295827). Furthermore, the melanoma score (initially termed the *PD-L1 Allred proportion score*[9] in earlier publications and abstracts[10–12]) was also based on a mixture of immune and tumor cell expression.[13] The use of the modified proportion and melanoma scores in those studies enabled further investigation of the role of PD-L1 expression in tumor and immune cells.

Initially, the mononuclear immune cell density score (MIDS), defined as the ratio of the number of PD-L1–expressing immune cells to that of all tumor cells, was used in an attempt to capture immune cell expression.[7] However, MIDS demonstrated poor reproducibility in preliminary studies. Given that the TPS is the ratio of the number of PD-L1–expressing tumor cells to that of all tumor cells, it is mathematically feasible to combine it with the MIDS (both are fractions with a common denominator). The result of this combination—the combined positive score (CPS)—is then, the ratio of the number of all PD-L1–expressing cells (tumor cells, lymphocytes, macrophages) to the number of all tumor cells. As the name implies, CPS considers PD-L1 expression on tumor cells and immune cells combined, the latter being a necessary inclusion if PD-L1 is to be a useful predictive biomarker for indications beyond NSCLC. In addition, CPS can be scored directly and reproducibly by a pathologist in one review of the slide and on a scale with similar resolution as that of TPS (ie, in deciles), obviating the need to score TPS and MIDS individually. This new scoring method, the PD-L1 CPS, is introduced here, along with the early work that led to its development and clinical validation and the evidence supporting its FDA approval as a companion diagnostic for identifying patients with gastric and gastroesophageal junction adenocarcinoma (hereafter, described as *gastric cancer*) who are most likely to respond to pembrolizumab therapy after prior treatment.

MATERIALS AND METHODS

Immunohistochemical staining for PD-L1 was performed on formalin-fixed, paraffin-embedded tumor samples from previously treated patients enrolled in the KEYNOTE-012, KEYNOTE-028, and KEYNOTE-059 studies, which used a version of the PD-L1 IHC 22C3 pharmDx assay labeled for investigational use only on the Autostainer Link 48 (Agilent), per the manufacturer's instructions,[14] as described in detail elsewhere.[15,16] PD-L1 expression was measured by 3 different methods: the CPS, the TPS, and the MIDS.

Scoring Methods

The CPS is given by summing the number of PD-L1–stained cells (tumor cells, lymphocytes, macrophages) and dividing the result by the total number of viable tumor cells, multiplied by 100, and is expressed by the following formula:

$$\mathrm{CPS} = \frac{\text{No. PD-L1–stained cells (tumor cells, lymphocytes, macrophages)}}{\text{Total No. of viable tumor cells}} \times 100$$

Although theoretically that quantity can exceed 100, the maximum score is defined as 100. A minimum of 100 viable tumor cells must be present in the PD-L1–stained slide (sectioned tumor biopsy or resection tissue) for the specimen to be considered adequate for evaluation. Tumor cells must show partial or complete membrane staining ($\geq 1^+$) to be counted as "stained," whereas immune cells are counted if there is any staining. For KEYNOTE-059 and reproducibility samples, results were prospectively considered PD-L1$^-$ if CPS was less than 1 and PD-L1$^+$ if CPS was 1 or more. Gastric cancer specimens stained for PD-L1 with this assay and scored as a CPS of less than 1 and 1 or more are shown in Figure 1, A and B, respectively.

The TPS is the percentage of viable tumor cells showing partial or complete membrane staining ($\geq 1^+$), relative to all viable tumor cells present in the sample (positive and negative) and is expressed by the following formula:

$$\mathrm{TPS}(\%) = \frac{\text{No. PD-L1–stained tumor cells}}{\text{Total No. of viable tumor cells}} \times 100$$

The MIDS is the ratio of PD-L1–stained immune cells to total number of viable tumor cells multiplied by 100. It is expressed by the following formula:

$$\mathrm{MIDS} = \frac{\text{No. PD-L1–stained immune cells}}{\text{Total No. of viable tumor cells}} \times 100$$

However, in practice, MIDS is scored on a scale of 0 to 4 because pathologists find it virtually impossible to score in deciles. The MIDS reflects the degree of PD-L1$^+$ immune cell infiltration into the tumor. The 5-point scale is somewhat similar to the "tumor infiltrating mononuclear" score described by Bellmunt et al,[17] although MIDS is defined more objectively. MIDS 0 means that there is no staining. MIDS 1 means that there is some staining below the threshold for MIDS 2. The threshold for a "positive" result is MIDS 2, which corresponds to 1 PD-L1$^+$ immune cell per 100 viable tumor cells. The thresholds for MIDS 3 and MIDS 4 are 10 and 100 PD-L1$^+$ immune cells per 100 viable tumor cells, respectively.

PD-L1 Staining in KEYNOTE-012 and KEYNOTE-028

KEYNOTE-012 (ClinicalTrials.gov, NCT01848834) was an open-label, phase 1b study of recurrent, metastatic, or persistent gastric or gastroesophageal cancer, urothelial cancer, triple-negative breast cancer, or squamous carcinoma of the head and neck, with any number of prior therapies. KEYNOTE-028 (ClinicalTrials.gov, NCT02054806) was an open-label, phase 1b study of 20 tumor types, including ovarian cancer, in which there was locally advanced and/or metastatic solid malignancy that was incurable and for which previous standard therapy was unsuccessful or was not appropriate. In both studies, objective response was assessed by reduction in tumor size (per Response Evaluation Criteria in Solid Tumors, version 1.1). All patients enrolled into KEYNOTE-012 and KEYNOTE-028 had tumors that were "positive" for PD-L1 (modified proportion score, ≥ 1) as determined by a prototype assay.[8] Tumor samples were subsequently retested using a version of the PD-L1 IHC 22C3 pharmDx assay labeled for investigational use only[15,16] and evaluated by an expert pathologist (blinded to clinical outcome and any prior descriptive work) to assess PD-L1 staining components of tumor cells and immune cells. Urothelial, gastric, triple-negative breast, and ovarian carcinomas, for which response data were available from KEYNOTE-012 and KEYNOTE-028 and for which initiation of registration trials were imminent, were selected for further development of scoring methods.

KEYNOTE-059

KEYNOTE-059 (ClinicalTrials.gov, NCT02335411) was a multicohort, open-label, phase 2 study of pembrolizumab alone or in combination with chemotherapy in recurrent or metastatic gastric or gastroesophageal cancer. KEYNOTE-059 cohort 1 included patients who received at least 2 prior chemotherapy regimens. Objective response was assessed by reduction in tumor size (Response Evaluation Criteria in Solid Tumors, version 1.1). Progression-free survival (PFS) and overall survival (OS) were also assessed using the Kaplan-Meier method. Tumor samples were tested with a version of the PD-L1 IHC 22C3 pharmDx assay labeled for investigational use only and were scored by pathologists who were not involved with the development of the assay but who were aware of the prespecified CPS cutoff of 1 or more.

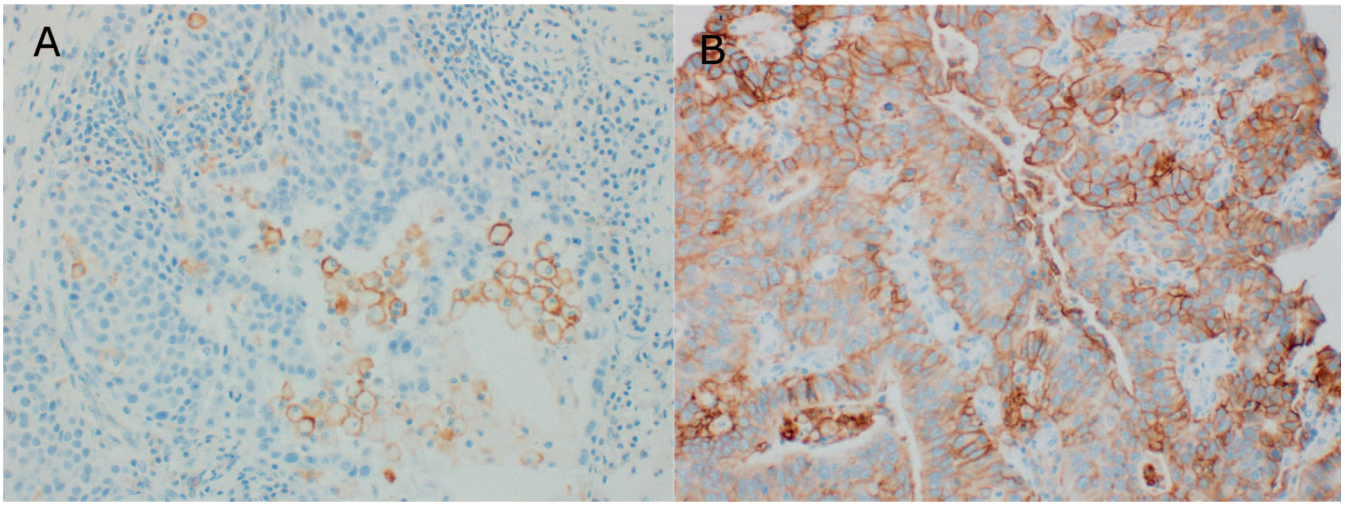

Figure 1. *Gastric cancer specimens stained with PD-L1 immunohistochemical 22C3 pharmDx (Dako North America, Agilent Technologies, Carpinteria, California). A, Specimen with CPS < 1. B, Specimen with CPS > 1. Abbreviations: CPS, combined positive score; PD-L1, programmed death ligand-1 (original magnification ×20 [A and B]).*

Testing and Scoring for Preliminary Reproducibility

Samples from KEYNOTE-012 were used to perform internal reproducibility studies for MIDS (3 pathologists), and then subsequently for CPS (2 pathologists). Pathologist 1 was an expert intimately involved with the assay development project. Pathologists 2 and 3 were otherwise uninvolved with development. All pathologists were blind to the clinical outcome and to any prior descriptive work. The MIDS 2 and CPS 1 were evaluated with 27 samples from the cohort with urothelial cancer. Results were summarized with an average negative agreement, an average positive agreement, and an overall agreement (OA).

Testing and Scoring for External Reproducibility

The investigational PD-L1 IHC 22C3 pharmDx assay was evaluated with respect to interpathologist and intrapathologist as well as intersite and intrasite reproducibility on formalin-fixed, paraffin-embedded human gastric cancer tissue specimens with a binary positive/negative CPS cutoff of 1 or more for determining PD-L1$^+$ and PD-L1$^-$ status.

Interpathologist and intrapathologist reproducibility was assessed with a set of 68 gastric cancer samples (34 PD-L1$^-$ and 34 PD-L1$^+$ samples) from the Agilent tissue bank, with a range of PD-L1 expression. Those samples were prestained with PD-L1 IHC 22C3 pharmDx at the Agilent development laboratory and sent to 3 external laboratories for evaluation. They were then scored by pathologists from each respective laboratory, none of whom were involved in the development of the assay or scoring guidelines. Each pathologist scored each sample 3 times on 3 nonconsecutive days with an intervening 2-week washout period. Therefore, the analyses were performed on 612 observations (68 samples, 3 pathologists, 3 scorings each). Each sample was compared with the appropriate positive/negative consensus call. For interpathologist reproducibility, the consensus was the majority call for the sample across all 9 observations. For intrapathologist reproducibility, the consensus was the majority call across 3 observations for the sample-pathologist combination.

Intersite and intrasite reproducibility of the technical staining procedure was also assessed (acknowledging that interpathologist reproducibility is inevitably nested within such a study). Fifteen replicate, unstained slides from each of 36 gastric cancer samples (18 PD-L1$^-$ and 18 PD-L1$^+$ samples) with a range of PD-L1 expression were used for testing. One set of replicate slides was tested and evaluated on each of 5 nonconsecutive days by each of 3 sites. Therefore, the analyses were performed on 540 observations. For intersite and intrasite reproducibility, the consensus was the majority call for the sample across all 15 observations and across 5 observations for the sample-site combination, respectively.

Positive/negative calls were made with respect to CPS of 1 or more for all reproducibility analyses. Positive percentage of agreement (PPA) and negative percentage of agreement (NPA) were calculated with respect to the consensus call; the OA was also calculated. To challenge the CPS of 1 or more cutoff, samples in the CPS 0–10 range were overrepresented in the reproducibility studies. Three techniques were used to reduce the chance that pathologists would have recall bias: (1) replicate samples were given unique identifiers and randomly assigned differently for each evaluation, (2) there was a minimum of a 2-week washout period between each time that a pathologist was allowed to score replicate samples, and (3) unique, nonreplicate "wild card" samples were included with each set of replicates.

RESULTS

PD-L1 Expression Among Tumors With Objective Responses From KEYNOTE-012 and KEYNOTE-028

Table 1 summarizes the pattern of PD-L1 expression in tumors that had an objective response to pembrolizumab from the cohorts with gastric, urothelial, triple-negative breast, and ovarian cancers in KEYNOTE-012 and KEYNOTE-028, according to PD-L1 positivity in the tumor and immune cells. Of 25 responders across all 4 tumor types, 12 (48%) expressed PD-L1 in the immune cell component alone and 1 (4%) expressed PD-L1 in the tumor cells alone, whereas 8 (32%) had PD-L1 expression in both the tumor and immune cell components. However, the combination of PD-L1 expression in either tumor cells, immune cells, or both resulted in the detection of 21 of the 25 responders (84%). In the patients with gastric cancer, immune-cell PD-L1 expression was observed in association with 9 of 11 responses (82%), with 2 of those 9 tumors (22%) expressing PD-L1 in tumor cells and immune cells, whereas tumor expression of PD-L1 alone was not associated with any of the responders.

Preliminary Reproducibility (CPS Versus MIDS)

Table 2 shows the results of the initial, preliminary reproducibility studies in urothelial carcinoma specimens.

Table 1. Pattern of Programmed Death Ligand-1 (PD-L1) Positivity in Tumor and Immune Cells Among Responders to Pembrolizumab From KEYNOTE-012 and KEYNOTE-028 Studies

Cancer Cohort (No. of Responders)	PD-L1+ Tumor Cells (TPS), %	PD-L1+ Immune Cells, No. (%)	
		<1:100	≥1:100
Gastric (n = 11)	<1	2 (18)	7 (64)
	≥1	0 (0)	2 (18)
Urothelial (n = 6)	<1	0 (0)	3 (50)
	≥1	0 (0)	3 (50)
Triple-negative breast (n = 5)	<1	1 (20)	1 (20)
	≥1	1 (20)	2 (40)
Ovarian (n = 3)	<1	1 (33)	1 (33)
	≥1	0 (0)	1 (33)
All 4 tumor types (n = 25)	<1	4 (16)	12 (48)
	≥1	1 (4)	8 (32)

Abbreviation: TPS, tumor proportion score.

The average negative agreement, average positive agreement, and OA of the 3 pathologists combined for MIDS were 47.1%, 60.0%, and 54.4%, respectively. Based on these initial data, the CPS scoring method was developed to simplify the assessment and inclusion of immune cells and to lead to better reproducibility. The average negative agreement, average positive agreement, and OA of 2 pathologists for CPS were 57.1%, 85.0%, and 77.8%, respectively. Based on these results, CPS was chosen for further development.

In KEYNOTE-012, 2, 4, and 5 gastric cancer samples had CPS scores of 0, 1–9, and 10 or more, respectively. Based on those results, a CPS of 1 or more was chosen as the cutoff for PD-L1 expression in gastric cancer. The CPS of 1 or more was equivalent to a TPS of 1% or greater or an MIDS of 1:100 or greater for those 11 samples. Although that equivalence is expected to hold for most samples, that is not a mathematical certitude. Samples can contain a mixture of low numbers of PD-L1–expressing tumor and immune cells that may reach the threshold of CPS of 1 or more but fall short of both TPS 1% and MIDS 1:100.

External Reproducibility

For interpathologist reproducibility of CPS, pairwise comparisons resulted in 591 concordant outcomes (313 negative and 278 positive), with a total of 21 discordant outcomes (10 discordant negative and 11 discordant positive). The NPA achieved 96.6% (313/324) agreement, the PPA achieved 96.5% (278/288) agreement, and the OA was 96.6% (591/612) (Table 3). For intrapathologist reproducibility, 612 total pairwise comparisons resulted in 595 concordant outcomes (315 negative and 280 positive), with a total of 17 discordant outcomes (8 discordant negative and 9 discordant positive). The NPA was 97.2% (315/324), the PPA was 97.2% (280/288), and the OA was 97.2% (595/612), with a lower bound of 95.3% (Table 3).

For intersite reproducibility of CPS, pairwise comparisons resulted in 497 total concordant outcomes (222 negative and 275 positive), with a total of 43 discordant outcomes (25 discordant negative and 18 discordant positive). The NPA and PPA were 92.5% (222/240) and 91.7% (275/300), respectively, and the OA was 92.0% (497/540) (Table 3). Intrasite reproducibility results were 93.1% (242/260), 98.2% (275/280), and 95.7% (517/540) for NPA, PPA, and OA, respectively.

Clinical End Points by CPS PD-L1 Expression

Objective response to pembrolizumab was significantly associated with CPS ($P = .002$, 1-sided rank-sum test) in KEYNOTE-059 (cohort 1), whereas the association with TPS was not significant ($P = .22$). Table 4 shows the objective response rates according to CPS and TPS. Based on a CPS cutoff of 1 or more, the prevalence of PD-L1 expression in this cohort was 57.6%, with reasonable enrichment of responses (odds ratio, 2.8). Based on a TPS cutoff of 1%, the prevalence of PD-L1 expression was 12.5%, with minimal enrichment (odds ratio, 1.4).

Figure 2, A, shows PFS, and Figure 2, B, shows OS by CPS PD-L1 expression in KEYNOTE-059 cohort 1. The PFS was significantly better in the PD-L1+ group than it was in the PD-L1− group, with a hazard ratio of 0.68 (95% CI, 0.52–0.88). The median PFS was 2.1 months (95% CI, 2.0–2.1 months) in the PD-L1+ group and 2.0 months (95% CI, 1.9–2.0 months) in the PD-L1− group. The OS showed a similar trend, although the difference did not reach statistical significance, with a hazard ratio of 0.76 (95% CI, 0.57–1.00). The median OS was 5.8 months (95% CI, 4.4–7.8 months) in the PD-L1+ group and 4.6 months (95% CI, 3.2–6.5 months) in the PD-L1− group.

DISCUSSION

This report introduces CPS as a new scoring method for assessing PD-L1 expression and describes the early work that led to the development of this scoring algorithm. Descriptive research in multiple tumor types from phase 1b

Table 2. Results of Preliminary Reproducibility Studies Performed on Urothelial Carcinoma Specimens[a]

Measure	Cutoff	Pathologists	Negative Pairs, No (%)	Discordant Pairs, No (%)	Positive Pairs, No (%)	Total No. of Pairs	Average Negative Agreement,[b] %	Average Positive Agreement,[c] %	Overall Agreement, %
CPS	≥1	1 and 2	4 (15)	6 (22)	17 (63)	27	57.1	85.0	77.8
MIDS	≥2	1 and 2	7 (28)	11 (44)	7 (28)	25[d]	56.0	56.0	56.0
		1 and 3	3 (11)	14 (52)	10 (37)	27	30.0	58.8	48.1
		2 and 3	6 (22)	11 (41)	10 (37)	27	52.2	64.5	59.3
		All 3	16 (20)	36 (46)	27 (34)	79	47.1	60.0	54.4

Abbreviations: CPS, combined positive score; MIDS, mononuclear immune cell density score.

[a] Studies were conducted internally at Agilent Technologies (Carpinteria, California).
[b] Average Negative Agreement = 2 × No. of Negative Pairs/[(2 × No. of Negative Pairs) + No. of Discordant Pairs].
[c] Average Positive Agreement = 2 × No. of Positive Pairs/[(2 x No. of Positive Pairs) + No. of Discordant Pairs].
[d] Excludes 2 specimens that both pathologists considered "not evaluable."

Table 3. Results From External Reproducibility Studies			
Reproducibility	Overall Percentage Agreement (95% CI)	Positive Percentage Agreement (95% CI)	Negative Percentage Agreement (95% CI)
Interpathologist	96.6 (94.0–98.7)	96.5 (93.1–99.3)	96.6 (92.9–99.4)
Intrapathologist	97.2 (95.3–98.9)	97.2 (94.8–99.3)	97.2 (94.8–99.1)
Intersite	92.0 (87.4–96.3)	91.7 (84.7–97.7)	92.5 (86.2–97.5)
Intrasite	95.7 (93.7–97.6)	98.2 (96.4–99.6)	93.1 (89.2–96.5)

study samples suggested that PD-L1 expression in tumor cells alone (TPS) would not be a useful predictive biomarker for response to pembrolizumab but that immune cell expression might be useful. Preliminary studies revealed that MIDS could not be scored reproducibly; therefore, the CPS was developed and showed more promise in terms of reproducibility. With refinement of scoring rules, the reproducibility of the CPS was confirmed. Finally, the clinical utility of CPS and its superiority over TPS in gastric cancer was confirmed in the KEYNOTE-059 study.

The clinical utility of CPS is underscored by the recent (September 2017) FDA approval of pembrolizumab for treatment of patients with gastric or gastroesophageal junction adenocarcinoma that expresses PD-L1.[1] Measurement of PD-L1 expression with the CPS in patients in KEYNOTE-059 cohort 1 reliably revealed a population of patients with gastric cancer likely to benefit from pembrolizumab who were otherwise undetected by TPS.[18] Although an "all-comers" approach is a common clinical strategy, the response to anti–PD-1/anti–PD-L1 therapy may be enhanced through PD-L1 evaluation of the combined tumor and immune cell components, potentially resulting in greater enrichment and higher response rates.

Although the enrichment of responders in KEYNOTE-059 was moderate (odds ratio, 2.8) compared with that reported previously for NSCLC (odds ratio, 8.9),[16] it was sufficient to identify a subgroup of patients likely to benefit from treatment and more than what was observed with TPS. Moreover, the objective response rate of 16.2% among patients with PD-L1$^+$ tumors using the CPS cutoff of 1 or more was sufficient to support FDA approval of pembrolizumab for this subgroup of patients. Additionally, PFS was significantly improved in the CPS PD-L1$^+$ compared with the PD-L1$^-$ population (hazard ratio, 0.68; 95% CI, 0.52–0.88) in patients in KEYNOTE-059 cohort 1; OS, although not statistically significant, confirmed that trend (hazard ratio, 0.76; 95% CI, 0.57–1.00). Although time-to-event analyses in single-arm cohorts should be viewed with caution, PFS and OS do corroborate the analyses of overall response rate. The data indicate that immune cell expression is a critical component in the CPS scoring algorithm for identifying patients with gastric cancers who are most likely to respond to therapy. High tumor mutational burden has also been identified as a predictor of sensitivity to anti–PD-1/anti–PD-L1 therapy across a range of tumor types.[19] Additional studies are necessary to evaluate the relative predictive values of PD-L1 expression with CPS and tumor mutational burden in gastric cancer and whether they can be used in conjunction to identify patients likely to benefit from pembrolizumab.

Although inclusion of immune cells in the scoring paradigm was expected to be more challenging than assessing tumor cells alone,[20,21] the CPS scoring method was shown to be highly reliable and reproducible. The reproducibility of MIDS did not seem promising in preliminary studies, but the reproducibility of CPS seemed reasonable, suggesting that CPS would be a better choice than MIDS for further development for indications in which immune cell expression is important for predicting the response to anti–PD-1/anti–PD-L1 therapy. In this report, external reproducibility was better than preliminary internal reproducibility, potentially resulting from a refinement in scoring rules. In fact, the reproducibility of CPS of 1 or more in gastric cancer seems to be comparable to, at a minimum, TPS 1% for NSCLC (albeit with slight methodological differences).[14]

Another advantage of CPS is that it eliminates the need to choose between tumor and immune cell PD-L1 expression as a predictive biomarker. Other schemes were considered for capturing tumor and immune cell PD-L1 expression in a single score. Perhaps the most obvious means was to use the percentage of total cells (tumor and immune cells) expressing PD-L1. However, such a score would decrease with increasing numbers of non–PD-L1–expressing, tumor-infiltrating immune cells, which most likely have antitumor activity. Results of several studies have shown that patients with tumors with more immune cell infiltration have better prognoses[22–25] and are more likely to respond to pembrolizumab.[26] The modified proportion score[8] and melanoma score[13] represent early efforts to incorporate immune and tumor cells into a single score. These scores are similar—the difference is that the modified proportion score is scored in deciles—whereas the melanoma score is based on larger "bins," originally defined by Allred et al.[9] In either case, how these scores are defined is awkward because the denominator included PD-L1–expressing immune cells only (excluding those without PD-L1 expression). In addition, those scores excluded stromal immune cells that were not adjacent to tumor cells,[8,13] which might have been ill-advised (tumor-infiltrating lymphocyte scores[27,28] and other PD-L1 scores[29,30] include all stromal immune cells). The CPS eliminates those weaknesses.

As with any IHC scoring method evaluated by a pathologist, subjectivity is an important consideration for the CPS. However, because the CPS is defined in terms of cell populations, it is versatile and well-suited for future

Table 4. Objective Response Rate (ORR) From KEYNOTE-059 by Combined Positive Score (CPS) and Tumor Proportion Score (TPS; N = 257)					
	Total (N = 257), No. (%)	Response	No Response	ORR, %	Odds Ratio
CPS ≥1	148 (57.6)	24	124	16.2	2.8
CPS <1	109 (42.4)	7	102	6.4	
TPS ≥1	32 (12.5)	5	27	15.6	1.4
TPS <1	225 (87.5)	26	199	11.6	

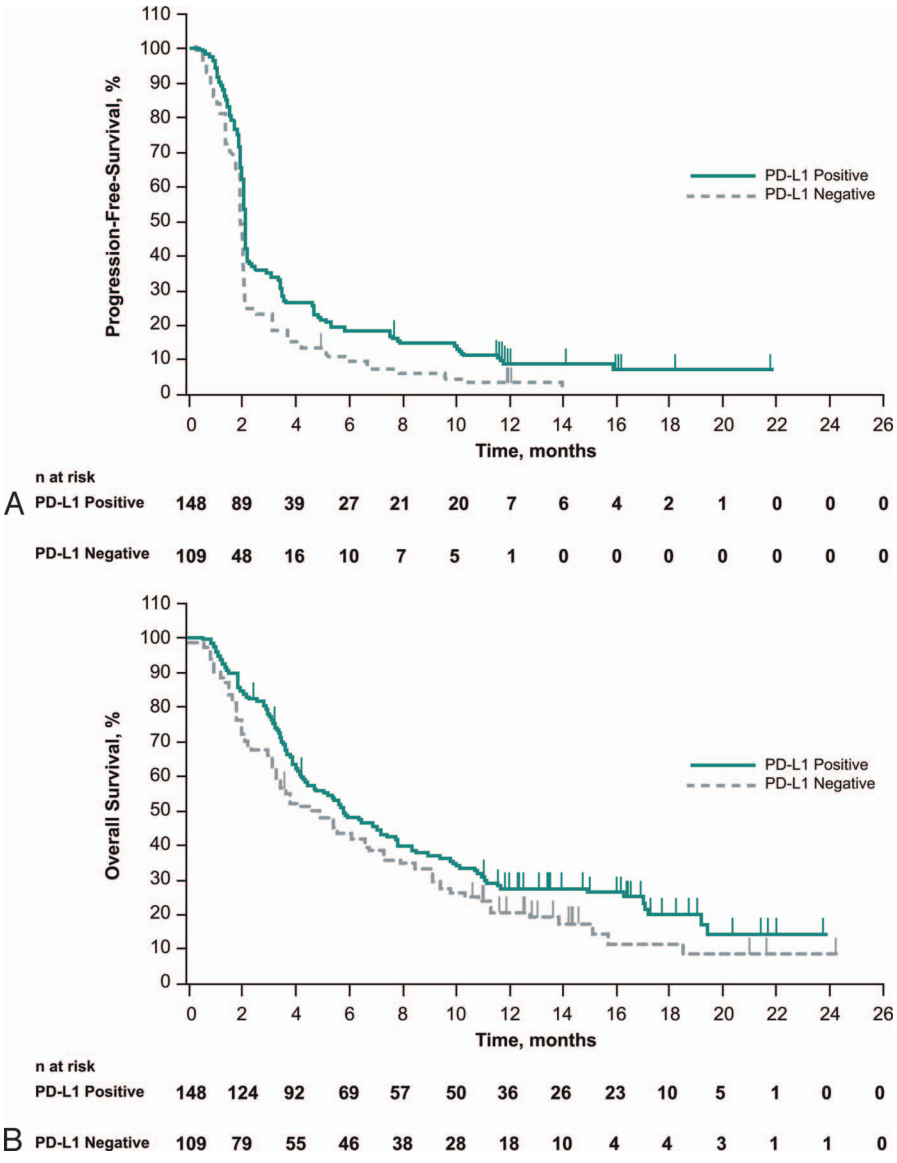

Figure 2. *A, Progression-free survival in patients from KEYNOTE-059 according to CPS PD-L1 expression. B, Overall survival. Abbreviations: CPS, combined positive score; PD-L1, programmed death ligand-1.*

applications; for example, the CPS can be applied to cytology specimens. This contrasts with immune-cell scoring methods based on the "percentage of tumor area occupied by any tumor-associated immune cells,"[29,30] which depend on tissue architecture. Theoretically, the CPS can also be used with flow cytometric methods, in which intact cells are recovered from a digested tumor. Finally, the CPS is amenable to automated scoring; if image analysis software can reliably recognize 3 cell populations (PD-L1–expressing tumor cells, PD-L1–expressing immune cells, and nonexpressing tumor cells), it can compute the CPS.

The CPS has been previously reported as a percentage,[31–33] although it is fundamentally a ratio of the number of PD-L1–expressing tumor cells, lymphocytes, and macrophages to the total number of viable tumor cells. To avoid expressing the CPS as a fraction, and by analogy with the TPS, the ratio was originally multiplied by 100%. Thus, for example, 1 PD-L1–expressing cell for every 100 tumor cells would be expressed as 1% instead of 0.01. This has generated a surprising amount of criticism, especially from regulators, suggesting that it is not appropriate to express the CPS as a percentage because the cells counted in the numerator are not a subset of those counted in the denominator. In response, the percent sign was eliminated, but the multiplication factor of 100 was retained. The correct interpretation of the CPS is the number of PD-L1–staining cells (both tumor and immune) for every 100 tumor cells. Notably, the numerical results are identical. Therefore, for example, the gastric cancer cutoff of CPS 1 is the same as that previously reported as CPS 1%.

One other minor debate during the development of this scoring method has been the formal definition of the CPS. The intent was to count all PD-L1–staining mononuclear immune cells in addition to membrane-staining tumor cells. Although the former technically includes monocytes and dendritic cells, these are rarely identified by pathologists when scoring PD-L1. Most tend to identify almost all large PD-L1–staining mononuclear immune cells as macrophages; thus, the parenthetic clarification "tumor cells, lymphocytes, and macrophages" in the formal CPS definition has proven to be an effective means of training pathologists.

Given that the relative importance of tumor versus immune cell PD-L1 expression seems to differ according to tumor type, it seems appropriate to reflect on whether a single underlying mechanism of action is operative in both cell types. Regarding tumor cell expression, the working hypothesis is that tumor cells express PD-L1 as a negative-feedback mechanism in response to cytokine release by effector immune cells. The question then arises whether that is also true for immune cells. Although it seems plausible that scavenger macrophages express PD-L1 in response to continued release of cytokines by effector T cells, it does not seem equally plausible that those same effector T cells express PD-L1 to modulate their own behavior. Another explanation might be that PD-L1–expressing immune cells have a regulatory function, similar to $FoxP3^+$ lymphocytes (tumor regulatory cells) and M2 macrophages. This seems to be fertile ground for future investigation.

Although the gastric cancer PD-L1 cutoff of CPS at 1 was prespecified for KEYNOTE-059, that cutoff was chosen based on limited data because of the rapid clinical development of pembrolizumab. KEYNOTE-012 was a biomarker-positive-only, indication-finding trial. Although it was not known for certain that PD-L1 expression would enrich for responders, the rationale was that, if it did enrich, it would be possible to demonstrate clinical activity more quickly and more economically than an all-comers study design. The disadvantage of this design is that it precluded evaluation of the false-positive rate, which is necessary for receiver operating characteristic analysis. Given that limitation, the approach was to choose a cutoff that would detect most responders and then evaluate that cutoff in the pivotal trial.

The CPS is a reproducible and versatile scoring method that enabled approval of pembrolizumab as a third-line treatment option for PD-L1–expressing gastric cancers, making pembrolizumab available to patients with few, if any, other treatment options. Together, pembrolizumab and PD-L1 IHC 22C3 pharmDx have the distinction of being the first 2 FDA-approved companion-diagnostic indications (NSCLC and gastric cancer) in cancer immunotherapy. The CPS incorporates PD-L1 expression by both tumor cells and immune cells into a single score that can be assessed directly by a pathologist, obviating the need to score tumor and immune cells separately. This makes the CPS useful in tumors in which PD-L1 expression is driven by tumor cells, immune cells, or both. Furthermore, the CPS is independent of tissue architecture, offering the promise of future utility in cytology or digested tumor samples. Therefore, the authors expect that the CPS will continue to be a useful diagnostic tool in immunotherapy.

Medical writing and/or editorial assistance was provided by Sarita S. Shaevitz, PhD, of the ApotheCom oncology team (Yardley, Pennsylvania). This assistance was funded by Merck & Co, Inc, Kenilworth, NJ.

References

1. Keytruda (pembrolizumab) for injection, for intravenous use [package insert]. Whitehouse Station, NJ: Merck Sharpe & Dohme Corp; 2017.

2. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012;366(26):2443–2454.

3. Garon EB, Rizvi NA, Hui R, et al; KEYNOTE-001 Investigators. Pembrolizumab for the treatment of non–small-cell lung cancer. *N Engl J Med*. 2015;372(21):2018–2028.

4. 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(7528):563–567.

5. Reck M, Rodriguez-Abreu D, Robinson AG, et al; KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823–1833.

6. Wolchok JD, Kluger H, Callahan MK, et al. Nivolumab plus ipilimumab in advanced melanoma. *N Engl J Med*. 2013;369(2):122–133.

7. Muro K, Chung HC, Shankaran V, et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet*. 2016;17(6):717–726.

8. Dolled-Filhart M, Locke D, Murphy T, et al. Development of a prototype immunohistochemistry assay to measure programmed death ligand-1 expression in tumor tissue. *Arch Pathol Lab Med*. 2016;140(11):1259–1266.

9. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998; 11(2):155–168.

10. Lote H, Cafferkey C, Chau I. PD-1 and PD-L1 blockade in gastrointestinal malignancies. *Cancer Treat Rev*. 2015;41(10):893–903.

11. Chen Z, Mei J, Liu L, et al. PD-L1 expression is associated with advanced non-small cell lung cancer. *Oncol Lett*. 2016;12(2):921–927.

12. Zhang P, Ma Y, Lv C, et al. Upregulation of programmed cell death ligand 1 promotes resistance response in non-small-cell lung cancer patients treated with neo-adjuvant chemotherapy. *Cancer Sci*. 2016;107(11):1563–1571.

13. Daud AI, Wolchok JD, Robert C, et al. Programmed death-ligand 1 expression and response to the anti-programmed death 1 antibody pembrolizumab in melanoma. *J Clin Oncol*. 2016;34(34):4102–4109.

14. Dako. PD-L1 IHC 22C3 pharmDx. SK006 50 tests for use with Autostainer Link 48. Carpinteria, CA: Dako North America Inc.; 2017. https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150013c.pdf. Accessed January 19, 2018.

15. 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(6): 392–397.

16. Dolled-Filhart M, Roach C, Toland G, et al. Development of a companion diagnostic for pembrolizumab in non–small cell lung cancer using immunohistochemistry for programmed death ligand-1. *Arch Pathol Lab Med*. 2016;140(11): 1259–1266.

17. Bellmunt J, Mullane SA, Werner L, et al. Association of PD-L1 expression on tumor-infiltrating mononuclear cells and overall survival in patients with urothelial carcinoma. *Ann Oncol*. 2015;26(4):812–817.

18. Kulangara K, Hanks DA, Waldroup S, et al. Development of the combined positive score (CPS) for the evaluation of PD-L1 in solid tumors with the immunohistochemistry assay PD-L1 IHC 22C3 pharmDx [ASCO abstract e14589]. *J Clin Oncol*. 2017;35.

19. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9(1):34.

20. 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*. 2016;12(2):208–222.

21. Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol*. 2016;29(10):1165–1172.

22. Thomas NE, Busam KJ, From L, et al. Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. *J Clin Oncol*. 2013;31(33):4252–4259.

23. Luen SJ, Salgado R, Fox S, et al. Tumour-infiltrating lymphocytes in advanced HER2-positive breast cancer treated with pertuzumab or placebo in addition to trastuzumab and docetaxel: a retrospective analysis of the CLEOPATRA study. *Lancet Oncol*. 2017;18(1):52–62.

24. Salgado R, Denkert C, Campbell C, et al. Tumor-infiltrating lymphocytes and associations with pathological complete response and event-free survival in HER2-positive early-stage breast cancer treated with lapatinib and trastuzumab: a secondary analysis of the NeoALTTO trial [published correction appears in *JAMA Oncol*. 2015;1(4):544; and *JAMA Oncol*. 2015;1(8):1172]. *JAMA Oncol*. 2015; 1(4):448–454.

25. Sudo T, Nishida R, Kawahara A, et al. Clinical impact of tumor-infiltrating lymphocytes in esophageal squamous cell carcinoma. *Ann Surg Oncol*. 2017; 24(12):3763–3770.

26. Loi S, Adams S, Schmid P, et al. Relationship between tumor infiltrating lymphocyte (TIL) levels and response to pembrolizumab (pembro) in metastatic triple-negative breast cancer (mTNBC): results from KEYNOTE-086 [ESMO abstract LBA13]. *Ann Oncol*. 2017;28(suppl 5).

27. Dieci MV, Radosevic-Robin N, Fineberg S, et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer [published online ahead of print October 9, 2017]. *Semin Cancer Biol*. 2017. doi: 10.1016/j.semcancer.2017.10.003.

28. Salgado R, Denkert C, Demaria S, et al; International TILs Working Group 2014. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer:

recommendations by an International TILs Working Group 2014. *Ann Oncol*. 2015;26(2):259–271.

29. Ventana PD-L1 (SP263) assay [package insert]. Tucson, AZ: Ventana Medical Systems, Inc; 2017.

30. Driver BR, Miller RA, Miller T, et al. Programmed death ligand-1 (PD-L1) Expression in either tumor cells or tumor-infiltrating immune cells correlates with solid and high-grade lung adenocarcinomas. *Arch Pathol Lab Med*. 2017; 141(11):1529–1532.

31. Bellmunt J, de Wit R, Vaughn DJ, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017;376(11):1015–1026.

32. Adams S, Loi S, Toppmeyer D, et al. Phase 2 study of pembrolizumab as first-line therapy for PD-L1–positive metastatic triple-negative breast cancer (mTNBC): preliminary data from KEYNOTE-086 cohort B [ASCO abstract 1088]. *J Clin Oncol*. 2017;35(15, suppl).

33. Keytruda 50 mg powder for concentrate for solution for infusion [summary of product characteristics]; Hoddesdon, UK: Merck Sharp & Dohme Limited; 2015. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003820/WC500190990.pdf. Accessed March 8, 2018.