Modern Pathology (2016) 29, 1165–1172

Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas

Andreas H Scheel[1], Manfred Dietel[2], Lukas C Heukamp[3], Korinna Jöhrens[2], Thomas Kirchner[4], Simone Reu[4], Josef Rüschoff[5], Hans-Ulrich Schildhaus[6], Peter Schirmacher[7], Markus Tiemann[3], Arne Warth[7], Wilko Weichert[8], Rieke N Fischer[9], Jürgen Wolf[9] and Reinhard Buettner[1]

[1]University Hospital Cologne, Institute of Pathology, Cologne, Germany; [2]Charité - University Hospital Berlin, Institute of Pathology, Berlin, Germany; [3]Institute for Hematopathology Hamburg, Hamburg, Germany; [4]LMU University Hospital Munich, Institute of Pathology LMU Munich, Munich, Germany; [5]Institute of Pathology Nordhessen, Kassel, Germany; [6]University Hospital Göttingen, Institute of Pathology, Göttingen, Germany; [7]University Hospital Heidelberg, Institute of Pathology, Heidelberg, Germany; [8]Technical University Munich (TUM), Institute of Pathology, Munich, Germany and [9]University Hospital Cologne, Medical Clinic I, Cologne, Germany

Immunohistochemistry of the PD-L1 protein may be predictive for anti-PD-1 and anti-PD-L1 immunotherapy in pulmonary adenocarcinoma and in clinically unselected cohorts of so-called non-small-cell lung cancer. Several PD-L1 immunohistochemistry assays with custom reagents and scoring-criteria are developed in parallel. Biomarker testing and clinical decision making would profit from harmonized PD-L1 diagnostics. To assess interobserver concordance and PD-L1 immunohistochemistry staining patterns, 15 pulmonary carcinoma resection specimens (adenocarcinoma: $n = 11$, squamous-cell carcinoma: $n = 4$) were centrally stained with the assays 28-8, 22C3, SP142, and SP263 according to clinical trial protocols. The slides were evaluated independently by nine pathologists. Proportions of PD-L1-positive carcinoma cells and immune cells were scored according to a 6-step system that integrates the criteria employed by the four PD-L1 immunohistochemistry assays. Proportion scoring of PD-L1-positive carcinoma cells showed moderate interobserver concordance coefficients for the 6-step scoring system (Light's kappa = 0.47–0.50). The integrated dichotomous proportion cut-offs (≥ 1, ≥ 5, ≥ 10, $\geq 50\%$) showed good concordance coefficients ($\kappa = 0.6$–0.8). Proportion scoring of PD-L1-positive immune cells yielded low interobserver concordance coefficients both for the 6-step-score ($\kappa < 0.2$) and the dichotomous cut-offs ($\kappa = 0.12$–0.25). The assays 28-8 and 22C3 stained similar proportions of carcinoma cells in 12 of 15 cases. SP142 stained fewer carcinoma cells compared to 28-8, 22C3, and SP263 in four cases, whereas SP263 stained more carcinoma cells in nine cases. SP142 and SP263 stained immune cells more intensely. The data indicate that carcinoma cells can be reproducibly scored in PD-L1 immunohistochemistry for pulmonary adenocarcinoma and squamous-cell carcinoma. No differences in interobserver concordance were noticed among the tested assays. The scoring of immune cells yielded low concordance rates and might require specific standardization. The four tested PD-L1 assays did not show comparable staining patterns in all cases. Thus, studies that correlate staining patterns and response to immunotherapy are required to test the significance of the observed differences.
Modern Pathology (2016) **29,** 1165–1172; doi:10.1038/modpathol.2016.117; published online 8 July 2016

Immunotherapy for pulmonary carcinomas may cause strong and durable anti-tumoral immune responses that may significantly improve overall survival.[1–4] One key signaling pathway is the interaction of aberrantly expressed PD-L1 ligand on carcinoma cell and tumor-associated immune cells, and the PD-1 receptor found on immune effector cells, most notably T-cells.[5] At least five therapeutic monoclonal antibodies against PD-1 or PD-L1 are clinically tested (Table 1A).[6] Immunohistochemistry of PD-L1 protein may be predictive for both kinds of inhibitor.[4,7–9] Each clinical trial is evaluating its own PD-L1 immunohistochemistry assay including

Correspondence: Dr AH Scheel, MD, Institute of Pathology University Hospital Cologne University of Cologne, Kerpener Street 62, Cologne 50937, Germany.
E-mail: andreas.scheel@uk-koeln.de
Received 6 March 2016; revised 9 May 2016; accepted 18 May 2016; published online 8 July 2016

Table 1 PD-L1 scoring-criteria

	Assay, antibody	*Cell type*	*Negative*	*Low/weak*	*Medium*	*High/strong*
A						
Nivolumab (α-PD-1; BMS)	Dako 28-8	Tumor	0–1%	1–5%	5–10%	≥10%
Pembrolizumab (α-PD-1; MSD)	Dako 22C3	Tumor	0–1%		1–50%	≥50%
Atezolizumab (α-PD-L1; Roche)	Ventana SP142	Tumor	0–1%	1–5%	5–50%	≥50%
		Immune	0–1%	1–5%	5–10%	≥10%
Durvalumab (α-PD-L1; AstraZeneca)	Ventana SP263	Tumor		1–25%		≥25%
Avelumab (α-PD-L1; Pfizer + Merck)	Dako	Tumor	0–1%		?	

		Negative		*Positive*			
B							
Proportion Score ('Cologne Score')	Category:	0	1	2	3	4	5
	Cut-off:	<1%	≥1%	≥5%	≥10%	≥25%	≥50%
	Interval:	0–1%	≥1% <5%	≥5% <10%	≥10% <25%	≥25% <50%	≥50% <75%

(A) At least five different monoclonal antibodies against PD-1 or PD-L1 are clinically tested in pulmonary carcinomas. For each therapeutic antibody an own PD-L1 immunohistochemistry assay is evaluated as predictive biomarker. Each assay includes its own primary antibody, detection system, and scoring-criteria. (B) The cut-offs of the different scoring-criteria may be integrated into a 6-step scoring system ('Cologne Score').

custom primary antibodies, staining platforms, and scoring-criteria[6] (Figure 1). Although some clinical trials specifically recruited patients with so-called 'squamous-cell non-small cell lung cancer'[1] and 'nonsquamous cell non-small-cell lung cancer,'[2] other trials recruited non-small-cell lung cancer patients independent of the respective histological type.[3,4,8,9]

Clinical decision making would profit from one harmonized procedure for PD-L1 testing in pulmonary adenocarcinoma and squamous-cell carcinoma. As a first step to gather experience with the clinical trial assays and to test the reproducibility of an integrated scoring system (Table 1B), we conducted a round robin test with two sets of centrally stained immunohistological specimens that were scored independently by nine pathologists. Two laboratory developed assays and four clinical trial assays stained according to the manufacturers' protocols were compared.

Materials and methods

Study Design

A scoring system was defined that combines all clinically tested proportions cut-offs (Table 1). The scoring system was tested on a set of $n=15$ pulmonary carcinoma resection specimens (adenocarcinoma $n=7$, squamous-cell carcinoma $n=8$) that were centrally stained for PD-L1 with two laboratory developed assays and scored by nine pathologists ('training set'). The data were validated by second round of $n=15$ pulmonary carcinoma resection specimens (adenocarcinoma $n=11$, squamous-cell carcinoma $n=4$) that were centrally stained for PD-L1 with four clinical trial assays ('validation set'). Each pathologist scored the same stained glass slides independently and blinded for the staining protocols. Proportion scores of carcinoma cells and of tumor-associated immune cells were recorded.

Cases

Specimens of pulmonary squamous-cell carcinoma and adenocarcinoma were selected from a comprehensively annotated patient cohort of a previous study.[10] The two sets were selected to include both histotypes to mimic the composition of clinical trials.[3,4,8,9] Permission of the institutional ethics committee was obtained. The histological types are listed in Supplementary Table 1.

Tissue Processing

Paraffin-sections of the formalin-fixed paraffin-embedded tissue were cut by two technicians in one session (training set) or by one technician in one session (validation set). Sections were mounted on positively charged, adhesive glass slides ('Clipped Corner X-tra Slides', Leica Biosystems, Wetzlar, Germany). Consecutive slides were used for the two stainings of the training set as well as for the four stainings of the validation set. Cut sections were stored and transferred at 2–6 °C and stained within 1 month.

Immunohistochemistry

Two laboratory developed assays for PD-L1 were set-up at the University Hospital Cologne, Institute of Pathology, using primary antibodies E1L3N[11] (Cell Signaling Technology, Cambridge, UK) and SP142 (ref. 7) (Spring Bioscience Corporation, Pleasanton, CA, USA) on an automated staining system with a polymer-based detection kit and DAB-chromogen (Leica Bond Polymer Refine; Leica Biosystems, Wetzlar, Germany).

Clinical trial assays were performed according to the manufacturers' instructions: Dako 28-8 pharmDx was stained at the Clinic for Dermatology, University Hospital Essen, Germany with a Dako certified

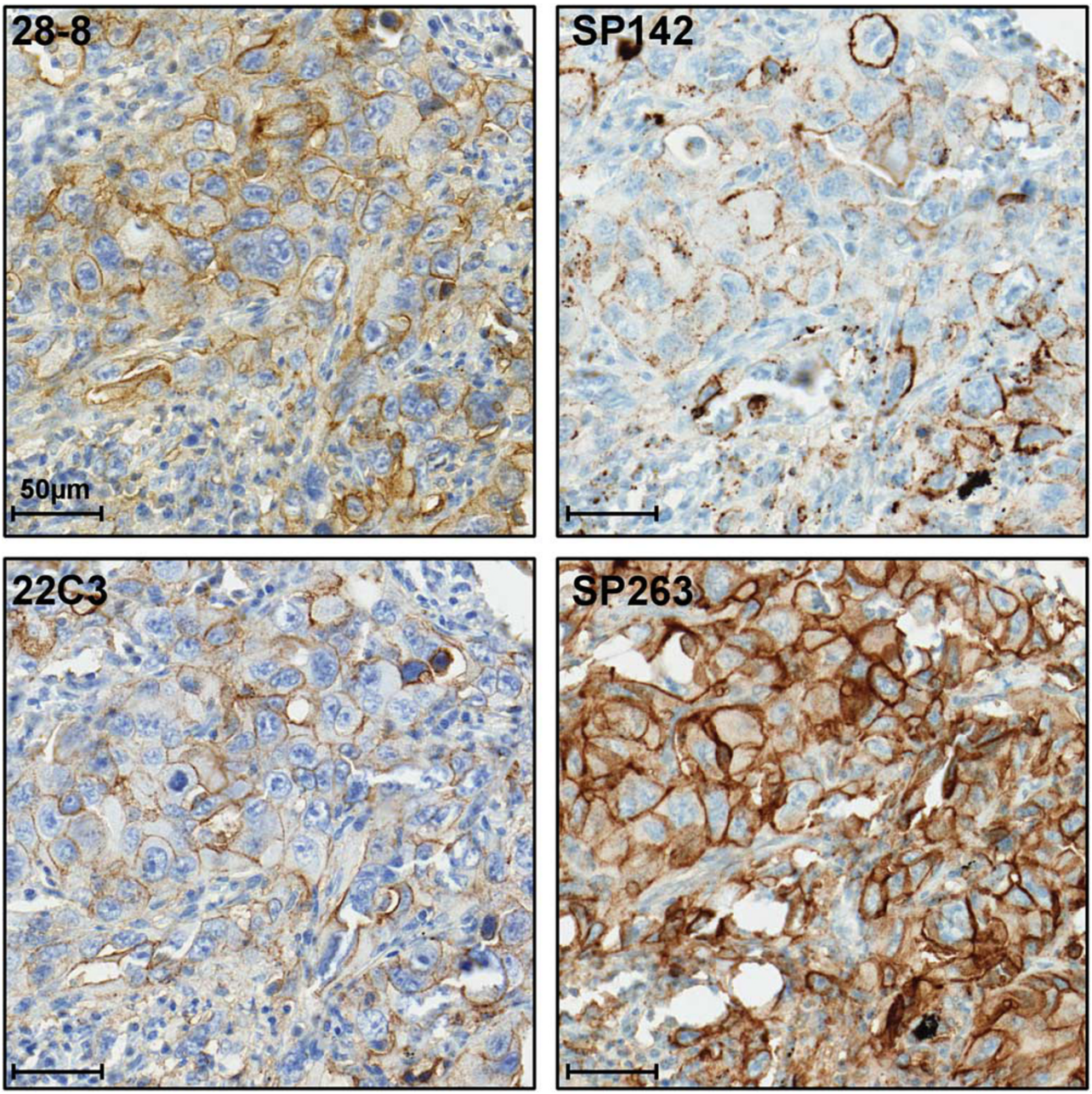

Figure 1 Staining patterns of clinical trial assays for PD-L1 immunohistochemistry. Example micrographs of four clinical trial assays for PD-L1 immunohistochemistry; matched regions on consecutive slides stained with the indicated assays. The case was scored PD-L1 positive (score 5, ≥50%) by all assays.

clinical trial set-up, Dako 22C3 pharmDx[8] was stained by Dako, Carpinteria, CA, USA. Dako assays were stained on the Dako Link AS-48 autostainer systems. The two Ventana assays SP142 and SP263 were stained by Ventana, Tucson, AZ, USA on Ventana Benchmark staining systems.

PD-L1 Immunohistochemistry Scoring

According to the de-facto consensus, a carcinoma cell was considered 'PD-L1 positive' if the cell membrane was partially or completely stained.[1–3,8,9] Cytoplasmatic PD-L1 staining in the carcinoma cells was disregarded. Carcinoma cells were quantified by evaluating the ratio of stained and unstained cells (Number of PD-L1-positive carcinoma cells÷Number of all carcinoma cells).[2,3]

An immune cell was considered 'PD-L1 positive' if it featured any PD-L1 staining (membrane/cytoplasm). PD-L1-positive immune cells are predominantly macrophages and lymphocytes. For lymphocytes, membranous and cytoplasmatic staining cannot be reliably distinguished due to the small

Table 2 Interobserver concordance for the scoring of PD-L1-positive carcinoma cells

	PD-L1 IHC	*Light's kappa (95% CI), tumor cell proportions*				
		6-step score	*Proportion cut-off*			
			≥1%	**≥5%**	**≥10%**	**≥50%**
Training set	E1L3N on Leica	0.50 [0.37–0.64]	0.73 [0.60–0.88]	0.79 [0.52–0.94]	0.74 [0.46–0.90]	0.76 [0.58–0.92]
Lab dev. assays	SP142 on Leica	0.49 [0.34–0.66]	0.61 [0.41–0.85]	0.77 [0.53–0.91]	0.79 [0.54–0.94]	0.80 [0.56–0.97]
$n=15$ cases NSCLC						
Validation set	Dako 28-8	0.49 [0.36–0.63]	0.79 [0.58–0.95]	0.63 [0.44–0.85]	0.65 [0.46–0.84]	0.77 [0.62–0.96]
Clinical trial assays	Dako 22C3	0.47 [0.34–0.63]	0.74 [0.44–0.94]	0.78 [0.58–0.94]	0.75 [0.52–0.89]	0.66 [0.42–0.89]
$n=15$ cases NSCLC	Ventana SP142	0.47 [0.35–0.62]	0.72 [0.53–0.89]	0.76 [0.49–0.93]	0.80 [0.62–0.93]	0.63 [0.46–0.80]
	Ventana SP263	0.47 [0.34–0.63]	0.59 [0.39–0.73]	0.78 [0.58–0.94]	0.77 [0.55–0.92]	0.75 [0.53–0.90]

Interobserver concordance of two sets of pulmonary squamous-cell and adenocarcinoma slides stained for PD-L1 with two laboratory developed assays ('training set') and four clinical trial assays ('validation set'). Light's Kappa (unweighted) and 95% CI for the 6-step score and 4 of the included cut-offs (square brackets) were calculated.

cell size. Immune cells were quantified by evaluating the ratio of the area covered by stained immune cells (Area covered by PD-L1-positive immune cells÷Tumor area) as has been described in studies with PD-L1 inhibitor atezolizumab.[4,9]

Necrotic areas were excluded from scoring. The commonly used minimum of ≥100 viable carcinoma cells was easily fulfilled in all cases given that the large resection specimens were used.

Integrated Proportion Score

At least five different PD-L1 immunohistochemistry assays are tested in clinical trials (Table 1A). Custom scoring-criteria are used for each assay. A common element of PD-L1 scoring is that the proportions of PD-L1-positive carcinoma cells (aka 'tumor cells') are estimated,[2,3] whereas staining intensity is not included.[7] For at least one assay (SP142), the area of PD-L1-positive immune cells is an independent part of the score.[4,9] The proportion cut-offs of the clinical trials overlap and may be integrated into a 6-step scoring system (Table 1B). The integrated score uses the categories 0–5 that match the clinical cut-offs. Thus, if a case is evaluated according to the scoring system, it can be classified by any of the included cut-offs.

Whole-Slide Scanning; Statistics

Micrographs were created by whole-slide scanning using a Pannoramic P250 scanner (3DHistech, Budapest, Hungary). Statistics were calculated with 'R' statistical programming language version 3.2.2 and package 'psy' version 1.1 by Bruno Falissard (www.r-project.org). Interobserver concordance was assessed using Cohen's kappa (unweighted) for two-rater-comparisons and Light's kappa (unweighted) for multirater comparisons.

Results

Scoring of Carcinoma Cells

For the training set of $n=15$ pulmonary carcinoma resection specimens (adenocarcinoma: $n=7$, squamous-cell carcinoma $n=8$; Supplementary Table 1), both laboratory developed assays yielded moderate concordance levels for the integrated proportion scoring of PD-L1 stained carcinoma cells (Table 2; Supplementary table 2): antibody E1L3N on Leica Bond staining system yielded a Light's kappa coefficient of $\kappa=0.50$ (95% confidence intervals (CI): 0.37–0.64), antibody SP142 on Leica Bond yielded a comparable Light's kappa of $\kappa=0.49$ (95% CI: 0.34–0.66). Classifying the cases by the dichotomous cut-off criteria that are included in the scoring system (≥1, ≥5, ≥10, ≥50%) resulted in good concordance levels of $\kappa=0.61$–0.80 (mean: 0.75).

For the validation set of $n=15$ pulmonary carcinoma resection specimens (adenocarcinoma: $n=11$, squamous-cell carcinoma: $n=4$), the four clinical trial assays yielded moderate concordance levels for the 6-step proportion score of stained carcinoma cells, $\kappa=0.47$–0.49 (Table 2; Supplementary Table 3). Classification by the included dichotomous cut-offs resulted in good concordance levels, $\kappa=0.59$–0.80 (mean: 0.72). No significant differences in concordance levels were noticed among the four assays.

To better quantify the interpretation-differences in the validation set, all 540 pairwise comparisons of the proportions-scores of two of the nine observers were plotted for each assay (15 specimens × C(9,2) observer-combination = 540 combinations per assay) (Figure 2). Similar frequencies of concordant pairs and of discordant pairs differing by one category or by ≥2 categories were noticed for each assay: 57–60% of pairs were concordant, whereas 25–32% differed by one category and 10–15% by two categories.

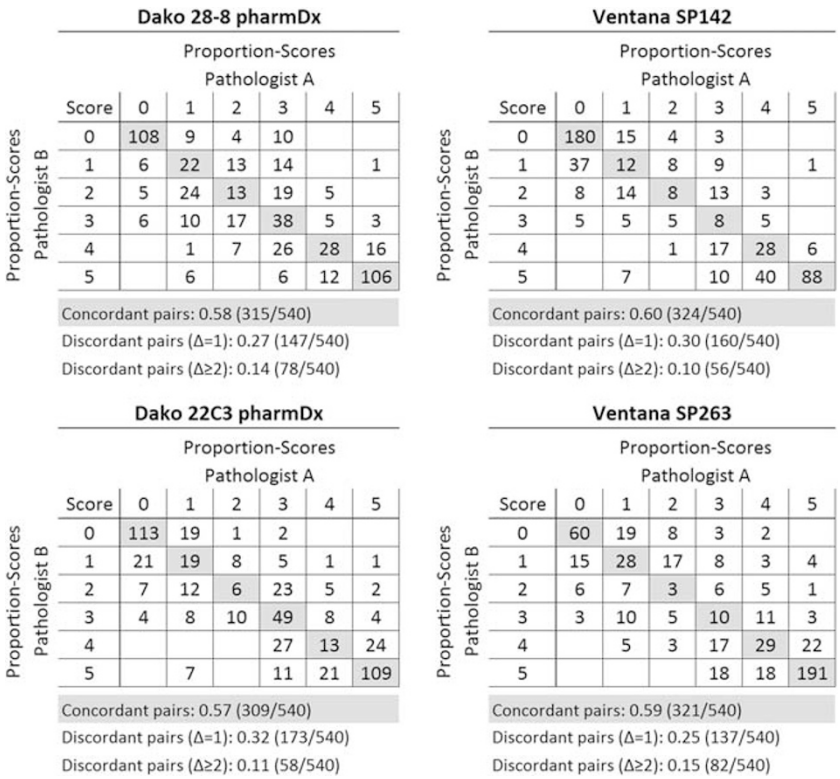

Figure 2 Interobserver concordance for the scoring of PD-L1-positive carcinoma cells. Pairwise comparisons of each sample and each combination of the nine observers (For each of the four assays: 15 specimens × C(9,2) observer-combination = 540 combinations). Each field indicates the absolute number of the respective score-pairing. Concordant scores (diagonal) are highlighted gray.

Scoring of Immune Cells

Scoring the area covered by immune cells relative to the tumor area[4,9] yielded low concordance levels in the training set and in the validation set for both, the integrated scoring system and the dichotomous cut-offs (mostly $\kappa < 0.2$) (Supplementary Table 4). It has to be noticed that 14 of the 15 cases of the validation set featured PD-L1-positive immune cells. The lack of cases clearly devoid of PD-L1-positive immune cells limits the interpretability of commonly used concordance coefficients[12] (Supplementary Table 2).

Assays

Scoring of the clinical trial assays indicated qualitative differences between the staining patterns: three assays showed a linear membranous staining of the carcinoma cells (28-8, 22C3, SP263), one assay showed a membranous, partially linear, partially granular staining (SP142) (Figure 1; Supplementary Figures 1–5). The differences in staining patterns were distinctive and enabled the observers to correctly sort the blinded validation set into three groups corresponding to the assays SP142, SP263, and 28-8/22C3 (data not shown).

The proportions of stained carcinoma cells appeared similar in most cases (Supplementary Table 3; Supplementary Figure 2). However, systematic differences were noticed, that particularly affected four cases: although two assays seemed comparable in intensity and proportions of stained tumor cells (28-8, 22C3), two assays appeared to be more intense (SP142, SP263) and to stain lower proportions (SP142) or higher proportions of carcinoma cells (SP263). Both SP142 and SP263 assays were noticed to stain immune cells more intense, which was especially evident in tonsil tissue used as control (Supplementary Figure 1).

To quantify potential differences in the proportions of stained carcinoma cells, the six pairwise comparisons of the four assays were plotted, each containing 135 data points (9 observer × 15 specimens) (Figure 3). For 28-8 *vs* 22C3, 72% of the pairs were concordant, 13% showed higher proportions for 28-8 and 16% higher proportions for 22C3. Comparisons of SP263 *vs* the other assays indicated higher proportions for SP263 in 46% (28-8), 44% (22C3), and 59% (SP142) of the pairs. Comparisons of SP142 indicated lower proportions in 36% (28-8), 39% (22C3), and 59% (SP263) of the pairs.

Conversely, different frequencies of the proportions were noticed among the assays (Figure 3, marginal sums): for 28-8 and 22C3, 24 and 26% of the 135 data points were negative (score 0), whereas 24 and 27% were strongly positive (score 5). For SP142 and SP263, 40 and 16% of the data points

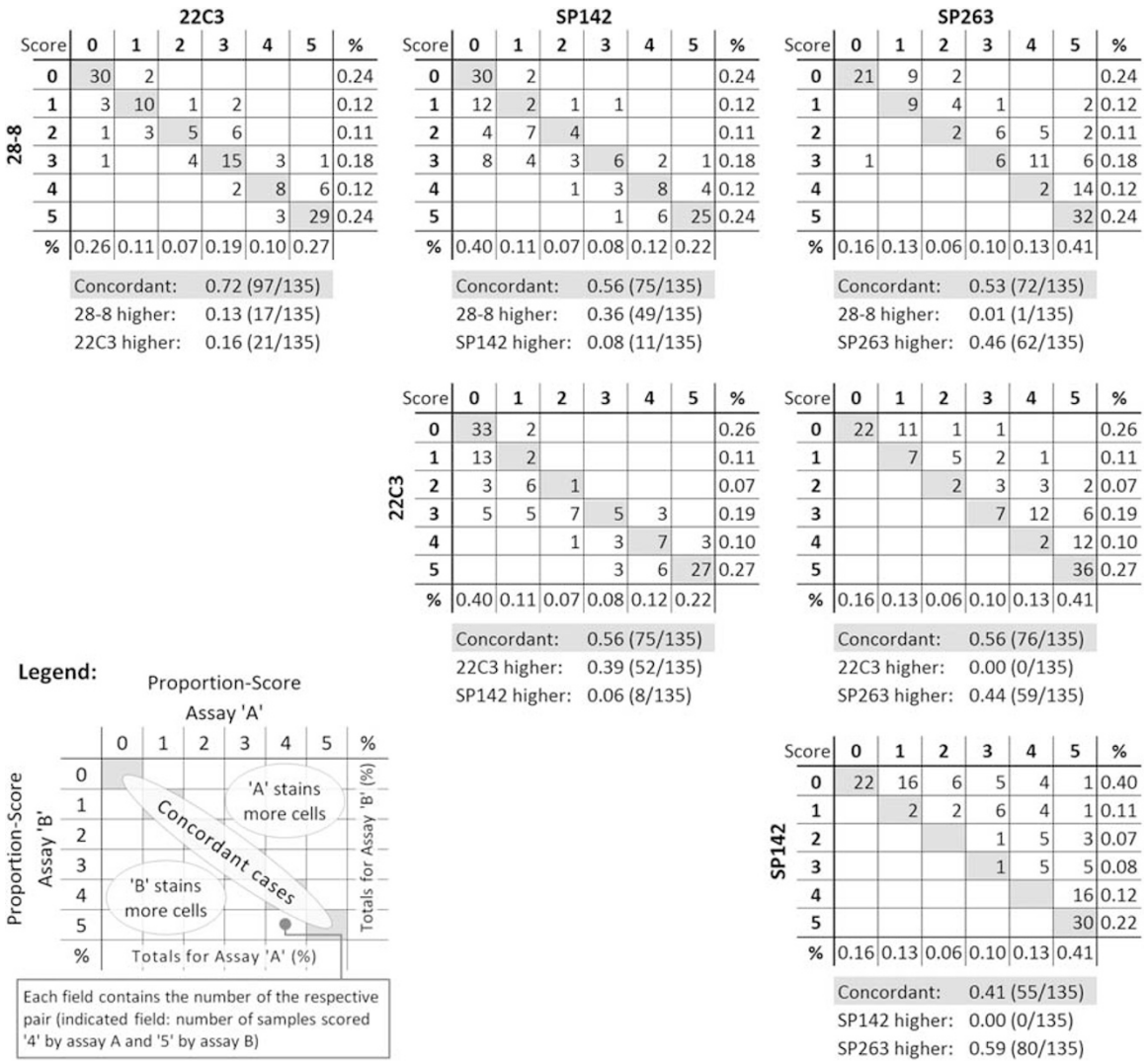

Figure 3 Pairwise comparisons of the four clinical trial assays. Analysis of the proportions of PD-L1-positive carcinoma cells stained by each assay. The six sub-tables show pairwise comparisons of two of the four clinical trial assays. For each comparison, 135 data points (15 samples × 9 observers) are mapped. Each field indicates the absolute number of the respective score-pairing. Concordant scores (diagonal) are highlighted gray. The marginal totals indicate the relative frequency of each scoring category for the respective assay.

were negative (score 0), whereas 22 and 41% were strongly positive (score 5).

Median scores of the nine observers were calculated for each sample and assay (Supplementary Table 5). By the median scores, assays 28-8 and 22C3 stained similar proportions of carcinoma cells in 12 of 15 cases. SP142 stained fewer carcinoma cells compared to 28-8, 22C3, and SP263 in four cases. SP263 stained more carcinoma cells in nine cases compared to the other assays. If the cases were to be classified according to a cut-off of $\geq 1\%$ stained carcinoma cells (Supplementary Table 6), 22C3 and 28-8 yielded similar results (11 of 15 positive cases). For SP142, two cases would have been classified negative that were classified positive by the other assays (9 of 15 positive cases) (Supplementary Figure 4). For SP263, two cases would have been classified positive that were scored negative by the other assays (13 of 15 cases positive). If a cut-off of $\geq 50\%$ stained carcinoma cells was applied (Supplementary Table 6), 22C3, 28-8, and SP142 yielded similar results (4 of 15 cases positive), whereas SP263 indicated two to be positive, that were $< 50\%$ with the other three assays (6 of 15 cases positive).

Given the low concordance rates for the scoring of immune cells, no cross-assay quantifications of the proportions and intensities of immune cells were made.

Discussion

In an early effort to harmonize PD-L1 immunohistochemistry in pulmonary squamous-cell and adenocarcinoma, two sets of $n = 15$ resection specimens were centrally stained for PD-L1 and scored by nine

pathologists. Scoring of the carcinoma cells was found to be reproducible for two laboratory developed assays and four clinical trial assays. Scoring of the tumor-associated immune cells yielded low concordance levels. The four clinical trial assays showed distinctive staining patterns and differences in the proportions of PD-L1 stained carcinoma cells.

The predictive value of clinical biomarkers relies on their biological significance and technical feasibility. For immunohistochemistry, the technical aspects encompass the employed assay and their reagents, the type and quality of the investigated biomaterial and the interpretation of the staining pattern by the pathologists.[13] Harmonization trials may greatly improve standardization and interobserver concordance for immunohistochemistry scoring as was demonstrated for HER2/neu[14] and ALK.[15] In this study, we focused on the scoring of PD-L1 immunohistochemistry to provide a starting point for the harmonization of the clinical assays (Table 1). The different cut-offs were integrated into a 6-step scoring system. In two sets of pulmonary carcinoma resection specimens including squamous-cell carcinoma ($n=12$) and adenocarcinoma ($n=18$), the score yielded moderate concordance levels ($\kappa\approx0.50$; PD-L1 stained carcinoma cells). The use of six scoring categories is a quite detailed way to interpret the rate of PD-L1-positive cells, in particular as the cut-offs 1, 5, and 10% are close to each other. Furthermore, large resection specimens were used, which complicated proportion scoring; thus, $\kappa\approx0.50$ seems an acceptable value. Moreover, the included cut-offs show good concordance levels of $\kappa=0.6$–0.8.

The integrated scoring system contains more information than a pure dichotomous cut-off. Several studies have indicated that the response rates to immunotherapy, median duration of responses, and overall survival times raise proportional to the rate of PD-L1-positive carcinoma cells,[2,3] indicating that the information might be clinically valuable.

Interestingly, no significant differences in the interobserver concordance levels were noticed among the four clinical trial assays. Apparently each assay can be interpreted reproducibly in pulmonary squamous-cell and adenocarcinoma, despite their differences in staining patterns. However, the finding cannot be directly translated to other entities: studies with HER2/neu indicated that specific guidelines and training are required for breast cancer and gastric cancer.[16] Analogous, entity-specific studies, and guidelines will be required for PD-L1 testing given that tumors with different morphologies and modes of carcinogenesis are investigated.[17,18]

Scoring of the tumor-associated immune cells yielded low concordance levels. The two sets were not perfectly suited for concordance analysis of immune cells as only two cases (training set) or one case (validation set) featured immune cells that stained PD-L1 negative by each assay. Such an extreme ratio of positive/negative cases limits the interpretability of concordance coefficients.[12] On the other hand, case discussions by the nine observers revealed that several aspects about immune cell scoring require standardization including a definition of 'tumor area' in not-well delimited carcinomas, a definition of area 'covered' by PD-L1-positive immune cells and a definition of the borders of necrotic areas. Given that the SP142 assay has been used reproducibly in published clinical trials,[4,9] we assume that specific instructions and training may raise concordance of immune cell scoring.

Although scoring of carcinoma cells seems to be reproducible for all tested assays, the staining results by the four clinical trial assays differed systematically: with the limitation of having seen only $n=15$ resection specimens, the assays yield differences in both the proportion of stained carcinoma cells as well as the staining intensity and proportion of stained tumor-associated immune cells. The findings are not surprising since each assay was developed and optimized as predictive biomarker for a different therapeutic antibody.[6] In particular, immune cells are part of the scoring for atezolizumab[8] (SP142 assay) but not for nivolumab[1,2] (28-8 assay), or pembrolizumab[3,8] (22C3 assay). Thus, it seems plausible that SP142 stains immune cells more intense. Our results give an early quantitation specifically why the clinical trial assays may not be used interchangeable.

A practical next steps towards harmonization could be to compare the staining patterns and predictive value of the 28-8 and 22C3 assays on a large cohort of samples. The limited number of cases investigated in our study indicates that the Dako assays might show comparable staining patterns in pulmonary squamous-cell carcinomas and adenocarcinomas. It would be of high interest to directly compare their predictive value.

Ultimately, the clinical utility of each PD-L1 immunohistochemistry assay has to be evaluated by its predictive value. An initial study design could be to stain PD-L1 in tumor samples of patients treated with immunotherapy with all four immunohistochemistry assays and compare their predictive value. Biochemical studies to determine the measurement range of each assay as well as mapping of the bound PD-L1 epitopes might facilitate interpretation of the immunohistochemistry results. If the divergent staining patterns that we report are confirmed by larger studies, relating PD-L1 immunohistochemistry of all assays to response and outcome of the same patient cohort will shed light on clinical significance of the staining differences.

Acknowledgments

We would like to express our gratitude to Bristol-Myers Squibb, MSD, Roche and AstraZeneca for providing us with the clinical trial assays. We would like to thank Prof. Dr D Schadendorf and A Sucker

for kindly performing the Dako 28-8 pharmDx stainings, and U Zenz, A Florin, and U Rommerscheidt-Fuß for excellent technical assistance. The work was supported by Bristol-Myers Squibb, MSD, Roche and AstraZeneca by providing the clinical trial assays and by supporting three joint meetings. The authors did not receive honoraria for conducting the harmonization-trial or for drafting the manuscript.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 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.
2 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.
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 Fehrenbacher L, Spira A, Ballinger M *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837–1846.
5 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;366:2443–2454.
6 Kerr KM, Tsao M-S, Nicholson AG *et al.* PD-L1 immunohistochemistry in lung cancer: in what state is this art?. J Thorac Oncol 2015;10:985–989.
7 Gettinger SN, Horn L, Gandhi L *et al.* Overall survival and long-term safety of nivolumab (anti-programmed death 1 antibody, BMS-936558, ONO-4538) in patients with previously treated advanced non-small-cell lung cancer. J Clin Oncol 2015;33:2004–2012.
8 Garon EB, Rizvi NA, Hui R. KEYNOTE-001 investigators: pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 2015;372:2018–2028.
9 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.
10 Seidel D, Zander T, Heukamp LC *et al.* A genomics-based classification of human lung tumors. Sci Transl Med 2013;5:209ra153.
11 Koh J, Go H, Keam B *et al.* Clinicopathologic analysis of programmed cell death-1 and programmed cell death-ligand 1 and 2 expressions in pulmonary adenocarcinoma: comparison with histology and driver oncogenic alteration status. Mod Pathol 2015;28:1154–1166.
12 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–268.
13 Bhargava R, Dabbs DJ. Immunhistology of the breast, theranostic applications. In: Dabbs DJ (ed). Diagnostic Immunohistochemistry: Theranostic and Genomic Applications, 4th edn. Elsevier Saunders: Philadelphia, PA, USA, 2014, pp 744–750.
14 Rüschoff J, Dietel M, Baretton G *et al.* HER2 diagnostics in gastric cancer-guideline validation and development of standardized immunohistochemical testing. Virchows Arch 2010;457:299–307.
15 Von Laffert M, Warth A, Penzel R *et al.* Multicenter immunohistochemical ALK-testing of non-small-cell lung cancer shows high concordance after harmonization of techniques and interpretation criteria. J Thorac Oncol 2014;9:1685–1692.
16 Rüschoff J, Hanna W, Bilous M *et al.* HER2 testing in gastric cancer: a practical approach. Mod Pathol 2012;25:637–650.
17 Robert C, Schachter J, Long GV. KEYNOTE-006 investigators: pembrolizumab versus ipilimumab in advanced melanoma. N Engl J Med 2015;372:2521–2532.
18 McDermott DF, Drake CG, Sznol M *et al.* Survival, durable response, and long-term safety in patients with previously treated advanced renal cell carcinoma receiving nivolumab. J Clin Oncol 2015;33:2013–2020.

Supplementary Information accompanies the paper on *Modern Pathology* website (http://www.nature.com/modpathol)