Modern Pathology (2021) 34:1125–1132
https://doi.org/10.1038/s41379-020-0644-7

ARTICLE

Comparison of three PD-L1 immunohistochemical assays in head and neck squamous cell carcinoma (HNSCC)

Emma J. de Ruiter[1] · Frans J. Mulder[2] · Bregje M. Koomen[1] · Ernst-Jan Speel[2] · Mari F. C. M. van den Hout[2] · Reinout H. de Roest[3] · Elisabeth Bloemena[4,5] · Lot A. Devriese[6] · Stefan M. Willems[1]

Received: 30 August 2019 / Revised: 22 July 2020 / Accepted: 23 July 2020 / Published online: 5 August 2020

Abstract

Expression of programmed cell death-ligand 1 (PD-L1) is being used as predictive biomarker for immunotherapy in head and neck squamous cell carcinoma (HNSCC). Several antibodies are available for PD-L1 testing and multiple staining and scoring methods are used. This study aimed to compare the performance of two PD-L1 standardized assays (SP263 and 22C3 pharmDx) and one laboratory-developed test (LDT) (22C3) in HNSCC using the tumor proportion score (TPS) and the combined positive score (CPS). Pretreatment biopsies from 147 HNSCC patients were collected in a tissue-microarray (TMA). Serial sections of the TMA were immunohistochemically stained for PD-L1 expression using 22C3 pharmDx on the Dako Link 48 platform, SP263 on the Ventana Benchmark Ultra platform, and 22C3 as an LDT on the Ventana Benchmark Ultra. Stained slides were assessed for TPS and CPS. Cutoffs of ≥1% and ≥50% for TPS and ≥1 and ≥20 for CPS were used. Concordance between the different staining assays was moderate to poor for TPS (intraclass correlation coefficient (ICC) 0.46) as well as for CPS (ICC 0.34). When stratifying patients by clinically relevant cutoffs, considerable differences between the assays were observed: concordance was poor for both TPS and CPS. Generally, SP263 stained a higher percentage of cells than the other assays, especially when using the CPS. Moderate concordance was shown between three different PD-L1 immunohistochemical assays and considerable differences in PD-L1 positivity were observed when using clinically relevant cutoffs. This should be taken into account when using PD-L1 expression to guide clinical practice.

Introduction

The development of immunotherapy with checkpoint-inhibitors such as monoclonal antibodies against programmed cell death protein 1 (PD-1) and programmed cell death-ligand 1 (PD-L1) has led to a significant improvement of treatment outcome in many types of cancer. In head and neck squamous cell carcinoma (HNSCC), PD-1 inhibitors pembrolizumab and nivolumab are approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for recurrent and metastatic disease and their implementation resulted in improved survival and reduced toxicity [1–5]. Combinations with chemotherapy and other PD-1/PD-L1 inhibitors atezolizumab, avelumab, cemiplimab, and durvalumab are currently being tested in clinical trials [6]. However, not every patient benefits from treatment with immune checkpoint inhibitors (ICI), as overall response rates range from 13–18% [2, 7].

Several studies have identified the expression of PD-L1 in tumor specimens as a predictive biomarker for treatment efficacy of PD-1 inhibitors. Currently, PD-L1 expression is

Supplementary information The online version of this article (https://doi.org/10.1038/s41379-020-0644-7) contains supplementary material, which is available to authorized users.

✉ Emma J. de Ruiter
e.j.deruiter@umcutrecht.nl

[1] Department of Pathology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

[2] Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands

[3] Department of Otolaryngology/Head and Neck Surgery, Amsterdam UMC, Cancer Center, Amsterdam, The Netherlands

[4] Department of Pathology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[5] Department of Maxillofacial Surgery/Oral Pathology, Amsterdam UMC and Academic Centre for Dentistry Amsterdam (ACTA), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[6] Department of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

being used as a selection marker for treatment with anti PD (-L)1 in lung cancer. In HNSCC, clinical trials evaluated the predictive value of PD-L1 expression as well. In the KEYNOTE-040 study, patients with recurrent or metastatic HNSCC treated with pembrolizumab showed a significantly improved survival when their tumor biopsies were positive for PD-L1, defined by a tumor proportion score (TPS) of ≥50% [3, 8]. For first-line treatment, PD-L1 expression performed most effectively as a predictive biomarker when using the combined positive score (CPS) with a cutoff of ≥20 (KEYNOTE-048) [6]. Other studies consider a cutoff of ≥1% for both TPS and CPS as clinically relevant as well [9]. Since June 2019, the FDA approval for pembrolizumab as first-line treatment of recurrent and metastatic HNSCC includes a CPS of ≥1 as selection criterion. Therefore, a reproducible and robust assay to quantify PD-L1 expression with comparable performance to the assay used in KEYNOTE -048 will have major clinical importance.

At this moment, several PD-L1 immunohistochemical assays are available. Most of them have been developed as a companion diagnostic test for treatment in clinical trials. The assays use different primary antibodies and staining platforms, as well as different ways of scoring and different clinically relevant cutoffs [10]. Besides the standardized assays, some laboratories have their own lab-developed tests (LDT). Two commonly used assays are Dako's 22C3 assay and Ventana's SP263 assay. The 22C3 assay runs on the Dako AS Link 48 IHC platform and is developed as a selection marker for pembrolizumab, while the SP263 assay runs on the Ventana Benchmark Ultra staining platform and was developed alongside durvalumab.

For other cancer types, studies showed a high concordance between these assays, which suggests that the assays might be used interchangeably [11, 12]. This would be highly valuable for patient care, as automated staining platforms are not universally available, causing a delay in diagnostics, and standardized assays are expensive. In HNSCC, only a few studies investigated the concordance between different PD-L1 immunohistochemical assays, with variable results [13–15].

This study aimed to compare the performance of the 22C3 and SP263 standardized assays, and an LDT using the 22C3 antibody for PD-L1 staining in HNSCC using both TPS and the CPS.

Methods

Patients and tumor specimens

This study was conducted using a consecutive, retrospective cohort of patients with HNSCC treated at the Amsterdam UMC (location VUmc) and the Maastricht University Medical Center, between January 2009 and December 2014. The patient cohort consisted of stage III or IV, HPV-negative oropharyngeal, hypopharyngeal, and laryngeal squamous cell carcinoma patients treated with radiotherapy with concomitant cisplatin or carboplatin with curative intent.

TMA construction

From all included patients, formalin fixed, paraffin embedded (FFPE) pretreatment biopsies were collected. Sections of the FFPE blocks were stained with hematoxylin and eosin, and assessed by a dedicated head and neck pathologist (SW, MvdH) to mark representative tumor regions. For each patient, three 0.6 mm tissue cores were obtained from the assigned area of the FFPE blocks and collected in a tissue-microarray (TMA). The TMA was constructed by a fully automated TMA instrument, as described before [16].

Immunohistochemistry

Serial sections of the TMA were immunohistochemically stained for PD-L1 expression using the standardized 22C3 pharmDx assay on the Dako Link 48 platform (Dako, Carpinteria, Ca), the standardized SP263 assay on the Ventana Benchmark Ultra platform (Ventana Medical Systems, Tucson, AZ), and 22C3 as an LDT on the Ventana Benchmark Ultra (dilution 1:80). The first assay has been used as standard in the KEYNOTE-048.

Pathological assessment of PD-L1 staining

The stained slides were simultaneously assessed by a dedicated head and neck pathologist, certified for PD-L1 testing, and a head and neck researcher; discrepancies were resolved by consensus (SW, EdR). Stainings were assessed for TPS and CPS. The TPS was defined as the number of positive tumor cells divided by the total number of viable tumor cells multiplied by 100%; the CPS as the number of positive tumor cells, lymphocytes and macrophages, divided by the total number of viable tumor cells multiplied by 100 (Fig. 1). Clinically relevant cut-offs of ≥1 and ≥50% for TPS and ≥1 and ≥20 for CPS were used. TMA cores that contained <100 viable tumor cells were excluded.

Statistical analysis

To compare the clinical performance of the assays, intra-class correlation coefficients (ICC) were calculated using the continuous PD-L1 scores of each individual TMA core, which was calculated based on a single-rating ($k = 2$), absolute-agreement, 2-way mixed-effects model [17].

PD-L1 scores per patient were calculated by taking the mean of all TMA cores taken from the same patient. Subsequently, patients were stratified using the above mentioned

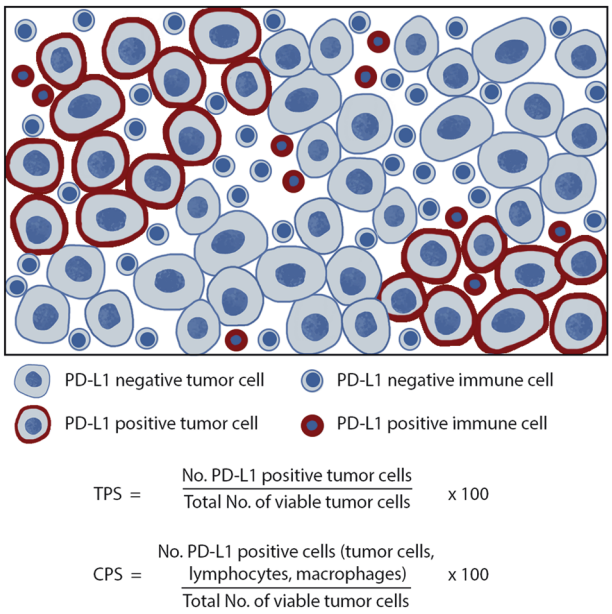

Fig. 1 Calculation of TPS and CPS. Schematic image of tumor specimen stained for PD-L1. Tumor proportion score (TPS) is defined as the number of positive tumor cells divided by the total number of viable tumor cells multiplied by 100%; combined positive score (CPS) as the number of positive tumor cells, lymphocytes and macrophages, divided by the total number of viable tumor cells multiplied by 100.

cutoffs and Cohen's kappas and confidence intervals (CI) were calculated. Overall percent agreement (OPA), positive percent agreement (PPA), and negative percent agreement (NPA) were calculated pairwise between the assays for all cutoffs; the 22C3 pharmDx assay was used as reference assay. For the comparison between the SP263 assay and the 22C3 LDT, the SP263 assay was used as reference assay.

To assess intratumor heterogeneity, weighted kappa was calculated between different TMA cores from the same patient. Interobserver variability between the two observers was calculated using ICC, calculated based on a single-rating ($k = 2$), absolute-agreement, 2-way mixed-effects model.

Of 12 randomly chosen tumor specimens, whole slides were stained for PD-L1 using the two standardized assays to assess the representativeness of the TMA cores. The concordance between the whole slides and the TMA cores was calculated based on a single-rating ($k = 2$), absolute-agreement, 2-way mixed-effects model.

Results

Patient and staining characteristics

A total of 147 head and neck tumors were eligible for inclusion. Patient characteristics are shown in Table 1.

Figure 2 shows representative images of a TMA core negatively stained for PD-L1 by the three assays (a–c) and a TMA core positively stained by the three assays (d–f). When using a cut-off of ≥50%, one tumor (0.7%) had a positive TPS in the 22C3 pharmDx assay, seven tumors (4.7%) in the SP263 assay, and five tumors (3.4%) in the 22C3 LDT. When using a cut-off of ≥1%, the TPS was positive in 40 (27.2%), 71 (48.3%), and 60 (40.8%) tumors, respectively. Regarding CPS, using a cut-off of ≥20 resulted in 9 positive tumors (6.1%) in the 22C3 pharmDx assay, 39 positive tumors (26.5%) in the SP263 assay, and 18 positive tumors (12.2%) in the 22C3 LDT. When using a cut-off of ≥1, the CPS was positive in 67 (45.6%), 128 (87.1%), and 94 (63.9%) tumors, respectively.

Table 1 Patient characteristics.

Age	58.7 (±6.4)	
Sex		
Male	99 (67.3%)	
Female	48 (32.7%)	
Tumor location		
Oropharynx	67 (45.6%)	
Hypopharynx	51 (34.7%)	
Larynx	29 (19.7%)	
T stage		
T1	4 (2.7%)	
T2	38 (25.9%)	
T3	44 (29.9%)	
T4a	42 (28.6%)	
T4b	19 (12.9%)	
N stage		
N0	27 (18.4)	
N1	21 (14.3%)	
N2a	12 (8.2%)	
N2b	43 (29.1%)	
N2c	37 (25.2%)	
N3	6 (4.1%)	
Nx	1 (0.7%)	
PD-L1 expression (TPS)	1–50%	≥50%
22C3 pharmDx	39 (26.4%)	1 (0.7%)
SP263	64 (43.2%)	7 (4.7%)
22C3 LDT	55 (37.2%)	5 (3.4%)
PD-L1 expression (CPS)	1–20	≥20
22C3 pharmDx	58 (39.2%)	9 (6.1%)
SP263	89 (60.5%)	39 (26.4%)
22C3 LDT	76 (51.7%)	18 (12.2%)

Comparison between the 22C3 pharmDx standardized assay and the SP263 standardized assay

When considering TPS, intraclass correlation between the 22C3 pharmDx and the SP263 assay was moderate (ICC

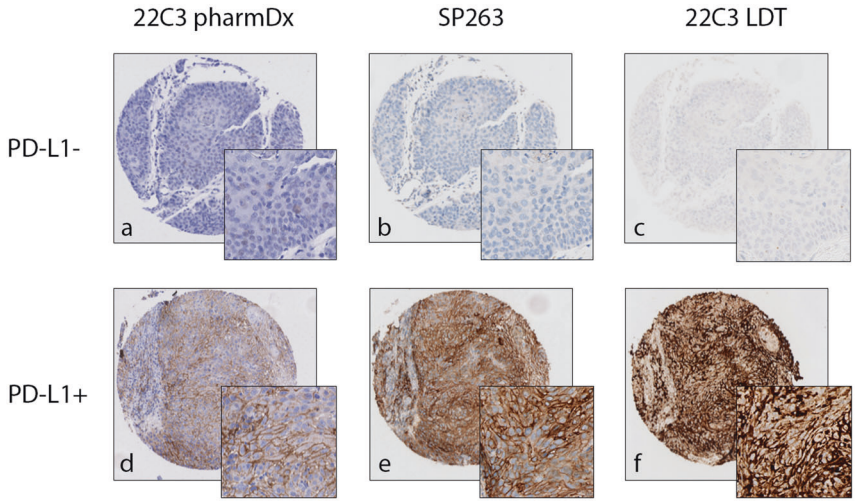

Fig. 2 Representative images of TMA cores. TMA cores negatively stained for PD-L1 using the 22C3 pharmDx assay (**a**), the SP263 assay (**b**), and a 22C3 LDT (**c**); TMA cores positively stained for PD-L1 using the 22C3 pharmDx assay (**d**), the SP263 assay (**e**), and a 22C3 LDT (**f**).

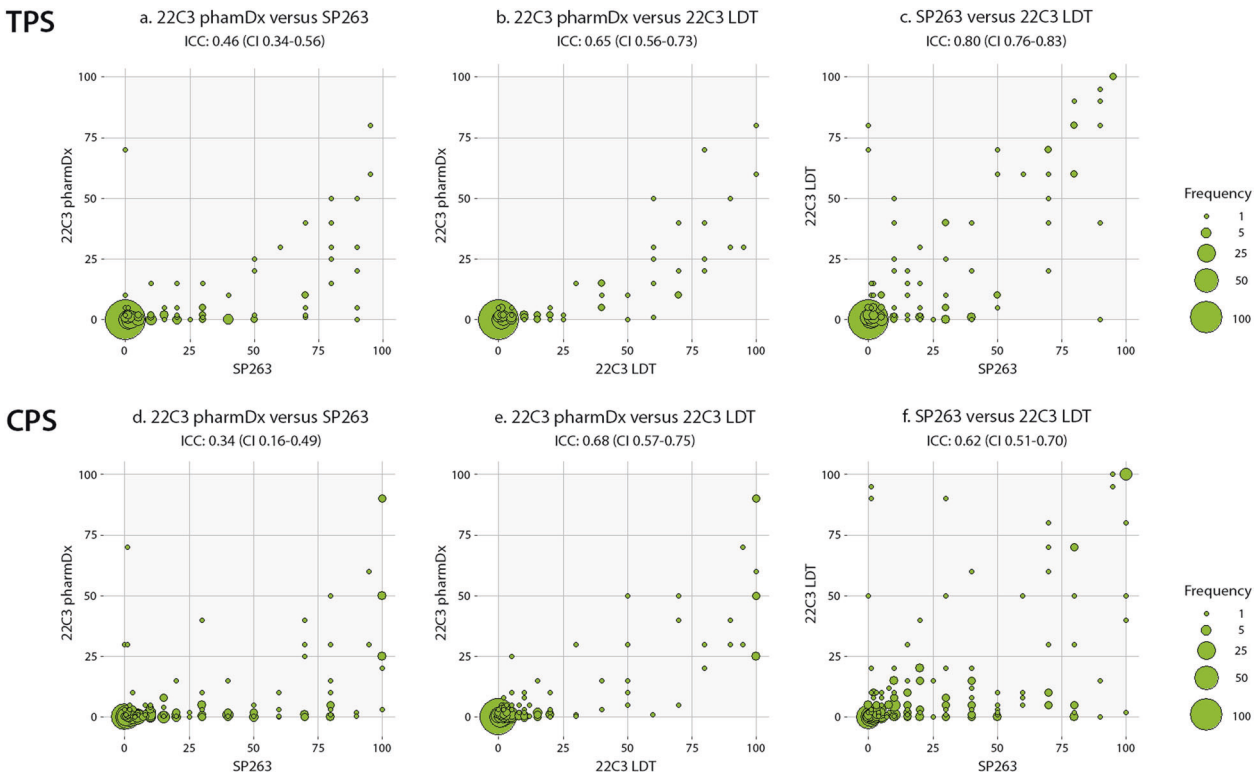

Fig. 3 Concordance between different PD-L1 immunohistochemical assays when scored on a continuous scale. a Concordance of TPS scores between 22C3 pharmDx and SP263. **b** Concordance of TPS scores between 22C3 pharmDx and 22C3 LDT. **c** Concordance of TPS scores between SP263 and 22C3 LDT. **d** Concordance of CPS scores between 22C3 pharmDx and SP263. **e** Concordance of CPS scores between 22C3 pharmDx and SP263. **f** Concordance of CPS scores between SP263 and 22C3 LDT.

0.46, CI 0.34–0.56) (Fig. 3a). However, after stratification of the tumors by a ≥ 50% cutoff, large differences between the two assays existed, with only one tumor testing positive when using the 22C3 pharmDx assay and seven tumors testing positive in the SP263 assay (kappa 0.24, CI 0–0.63) (Table 2). It should be noted, however, that statistics should be interpreted with caution due to the low percentage of PD-L1 positivity in this cohort. At a cutoff of ≥1%, concordance was moderate to poor (kappa 0.43, CI 0.30–0.57). Differences in PD-L1 positivity according to the TPS for individual tumors using different assays, including the 22C3 LDT, are visualized in Fig. 4a.

For CPS, even less concordance was observed between the SP263 and the 22C3 pharmDx assay. With an ICC of 0.34 (CI 0.16–0.49) and kappa values of 0.22 (CI 0.13–0.32) and 0.26 (CI 0.10–0.42) at a ≥ 1 and ≥20 cutoff,

respectively, concordance can be considered to be poor (Fig. 3d, Table 3). Differences in CPS for individual tumors using the three different assays are visualized in Fig. 4b.

OPA, PPA, and NPA values of the assays for each clinically relevant cutoff are shown in Supplementary Tables 1a, b.

Comparison between the two standardized assays and the 22C3 LDT

For TPS, the 22C3 LDT seemed to be more concordant with the SP263 assay (ICC 0.80, CI 0.76–0.83; ≥1% kappa 0.55, CI 0.41–0.68; ≥50% kappa 0.64, CI 0.42–0.85) than with the 22C3 pharmDx assay (ICC 0.65, CI 0.56–0.73; ≥1% kappa 0.47, CI 0.32–0.61; ≥50% kappa 0.33, CI 0–0.81). For CPS, this was the other way around, although both the concordance with the SP263 assay (ICC 0.62, CI 0.51–0.70); ≥1 kappa 0.21, CI 0.066–0.36; ≥20 kappa 0.47, CI 0.31–0.64) and with the 22C3 assay (ICC 0.68, CI 0.57–0.75; ≥1 kappa 0.43, CI 0.29–0.56; ≥20 kappa 0.64, CI 0.42–0.85) could be defined as moderate to poor (Fig. 3b, c, e, f, Table 3).

Intratumor heterogeneity

When considering TPS, concordance between TMA cores from the same patients was good for the SP263 assay and the 22C3 LDT, and moderate to good for the 22C3 pharmDx assay. For CPS, the intratumor heterogeneity was generally higher: concordance was moderate to good for the SP263 assay and the 22C3 LDT, and moderate for the 22C3 pharmDx assay. All kappa coefficients are shown in Table 4.

Interobserver variability and validity of the use of TMA cores

Although not the primary aim of this study, interobserver variability between the two observers was assessed. Good concordance was observed between the observers for all assays, especially for CPS (Supplementary Table 2).

Concordance between PD-L1 scores based on TMA cores and whole slides was generally good (Supplementary table 3). Concordance was better for TPS than for CPS. The observation that the SP263 assay results in higher TPS and CPS scores than the 22C3 pharmDx assay was supported by this analysis: when comparing the SP263 assay with the 22C3 pharmDx assay, five out of twelve tumors had a higher TPS and ten out of twelve had a higher CPS. None of the stained slides scored higher in the 22C3 assay (Supplementary figs. 1 and 2).

Table 2 Comparison of TPS between three PD-L1 immunohistochemical assays using cutoffs of ≥1% and ≥50%.

Cutoff: ≥1%	Kappa (CI)
22C3 pharmDx vs. SP263	0.43 (0.30–0.57)
22C3 pharmDx vs. 22C3 LDT	0.47 (0.32–0.61)
SP263 vs. 22C3 LDT	0.55 (0.41–0.68)
Cutoff: ≥50%	
22C3 pharmDx vs. SP263	0.24 (0–0.63)
22C3 pharmDx vs. 22C3 LDT	0.33 (0–0.81)
SP263 vs. 22C3 LDT	0.65 (0.33–0.97)

Table 3 Comparison of CPS between three PD-L1 immunohistochemical assays using cutoffs of ≥1 and ≥20.

Cutoff: ≥1	Kappa (CI)
22C3 pharmDx vs. SP263	0.22 (0.13–0.32)
22C3 pharmDx vs. 22C3 LDT	0.43 (0.29–0.56)
SP263 vs. 22C3 LDT	0.21 (0.066–0.36)
Cutoff: ≥20	
22C3 pharmDx vs. SP263	0.26 (0.10–0.42)
22C3 pharmDx vs. 22C3 LDT	0.64 (0.42–0.85)
SP263 vs. 22C3 LDT	0.47 (0.31–0.64)

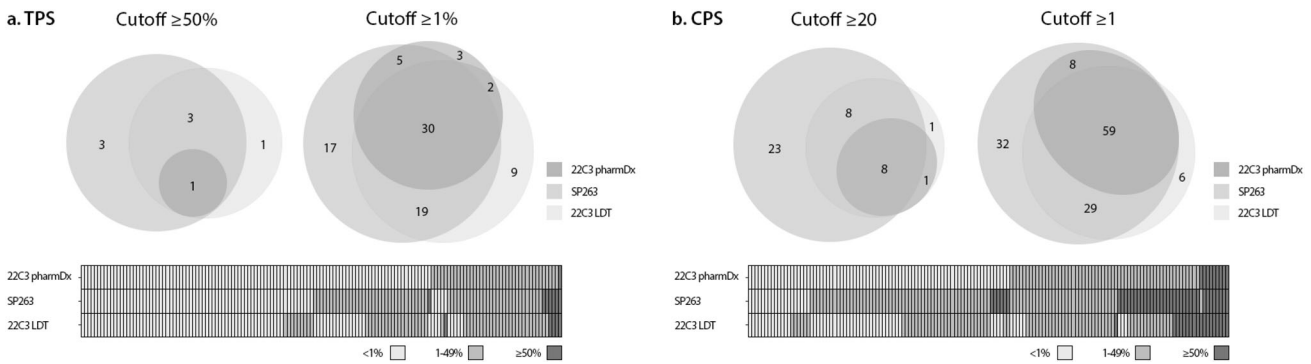

Fig. 4 Concordance between different PD-L1 immunohistochemical assays in the individual patient. a Venn diagrams of PD-L1 positivity using TPS cut-offs of ≥50% and ≥1%, and a heatmap visualizing differences in TPS within individual patients. **b** Venn diagrams of PD-L1 positivity using CPS cut-offs of ≥20 and ≥1, and a heatmap visualizing differences in CPS in individual patients.

Table 4 ICC between tumor cores of the same patient.

	TPS	CPS
22C3 pharmDx	ICC 0.79 (0.70–0.85)	ICC 0.64 (0.49–0.75)
SP263	ICC 0.85 (0.78–0.89)	ICC 0.70 (0.58–0.79)
22C3 LDT	ICC 0.85 (0.78–0.89)	ICC 0.82 (0.75–0.88)

Intraclass correlation coefficients (ICC) of PD-L1 expression between three PD-L1 immunohistochemical assays for TPS and CPS.

Discussion

The development of ICI has led to a revolution in the treatment of cancer. However, as not every patient benefits from this type of immunotherapy, predicting which patients are likely to respond is of major importance. PD-L1 expression, based on immunohistochemical evaluation on tumor cells and/or immune cells, is used as a selection marker for immunotherapy with ICI in several cancer types. In HNSCC, this is recently implemented for immunotherapy with pembrolizumab as first-line treatment of recurrent and metastatic HNSCC, facing pathology departments with an increasing demand for PD-L1 testing of tumor specimens.

Due to practical reasons, it is preferred to use different PD-L1 immunohistochemical assays interchangeably. Therefore, we aimed to investigate the concordance between two standardized PD-L1 assays (the 22C3 pharmDx and the SP263 assay) currently used in diagnostics and one LDT using the 22C3 antibody. Both TPS and CPS were assessed, using clinically relevant cutoffs of ≥1% and ≥50% for TPS, and ≥1 and ≥20 for CPS.

Our study suggests that considerable differences exist between the two standardized assays that are currently being used in diagnostics in different laboratories in The Netherlands. Intraclass correlation analyses on the continuous data showed moderate concordance between the antibodies, with the SP263 assay structurally resulting in a higher PD-L1 score than the 22C3 pharmDx assay, for TPS as well as CPS. More importantly, after stratification in positive-and negative-tumors based on the observed PD-L1 expression, significant differences were observed between the two assays, especially when using cutoffs of ≥20 and ≥50%, for CPS and TPS respectively.

These findings deviate from two other studies that assessed the concordance between the 22C3 pharmDx and the SP263 assay in HNSCC: Ratcliff et al. (2016) reported fair concordance between the two assays comparing PD-L1 expression in 108 HNSCC biopsy samples and suggested the possibility of using the assays interchangeably [15]. Hodgson et al. (2018) compared PD-L1 positivity in 27 surgically resected hypopharyngeal tumors and report a moderate to substantial concordance, with higher PD-L1 positivity rates using the SP263 assay [14]. In lung cancer, some studies did report similar results to our study: Munari et al. (2018) mention significantly less positivity when using the 22C3 assay compared with the SP263 assay and show important discrepancies in identifying positive cases at clinically relevant cutoffs [18]. Nevertheless, most studies showed fair concordance between the two assays, although conclusions on whether to use them interchangeably or not, differ.

Besides the two standardized assays, we also assessed the diagnostic performance of an LDT using the 22C3 antibody. Results of this LDT differed significantly from the two standardized assays, although it showed more concordance with the SP263 assay than with the 22C3 assay, especially for TPS. This is in contrast with the meta-analysis of Torlakovic et al., which concludes that a well-designed LDT may achieve higher accuracy than a standardized test designed and approved for a different purpose [12]. However, no studies assessing HNSCC were included in the meta-analysis and it is not unlikely that correlation in immunohistochemical expression using different antibodies might vary between tumor types, it should also be noted, that the LDT used in this study was developed in the diagnostic laboratory that is using the SP263 assay by default and this assay was also used for the optimization and validation of the LDT on the Ventana autostainer.

Our study identified only a small number of PD-L1 positive tumors in a relatively large cohort of 147 head and neck tumors when using clinically relevant cut-offs. Although some other studies report positivity rates below 5% [14], the percentage of tumors with a TPS above 50% or a CPS above 20 in our cohort according to the 22C3 pharmDx assay was 0.7% and 6.1%, respectively. Several reasons could underlie this difference with the literature. Firstly, we used a homogeneous patient cohort consisting of stage III and IV, HPV negative HNSCC patients without distant metastases scheduled for primary chemoradiation, while most studies use very heterogeneous patient cohorts. None of the included patients received any prior treatment, while most clinical trials concern recurrent or metastatic disease. Secondly, because all patients were scheduled for conservative treatment, only small biopsies were available for PD-L1 testing, so tumor heterogeneity might play a larger role than in studies using surgically resected tumor specimens. Thirdly, the low prevalence of PD-L1 positivity in our cohort might lie in the optimization of the immunohistochemical assays. However, the 22C3 pharmDx assay and SP263 assay used in this study are standardized tests and are both currently used in diagnostics. We therefore believe that the prevalence of positive cases and the corresponding comparison of the three assays in this study are a realistic reflection of PD-L1 testing in clinical diagnostic practice.

Another problem in the immunohistochemical evaluation of HNSCC is the high intratumor heterogeneity. Although

individual cores of the same patient generally showed a fair concordance in our study, intratumor heterogeneity was observed in a considerable number of tumors. This should be taken into account when using PD-L1 expression as a selection marker for immunotherapy with ICI, because biopsies of HNSCC are generally small, and by basing PD-L1 positivity on such a small amount of tumor tissue patients that could benefit from ICI might be missed. A study design comparing PD-L1 expression in tumor biopsies compared with surgical resections of the same tumor could give insight in how the pretreatment biopsy relates to the whole tumor. A study of Scott et al. (2017), for example, showed high inter- and intra-tumor block concordance on a small number of HNSCC tumors using the SP263 assay [19].

Our study had some limitations: firstly, TMA's were used instead of whole tissue slides. Three TMA cores were assessed for each tumor, which was considered representative for the tumor biopsies. In order to prevent differences in PD-L1 scoring due to tumor heterogeneity, serial sections of the same TMA cores were used to analyze concordance. The use of TMA's might result in deviant CPS scores, since the TMA's were not specifically constructed to assess the tumor microenvironment, which was supported by the finding that concordance between TMA and whole slides was better for TPS score than for CPS score.

Secondly, PD-L1 expression in tumor specimens was scored by only two observers and definitive scores were based on consensus. It is clear that interobserver variability plays an important role in diagnostics. The aim of this study was to compare the performance of different PD-L1 assays; assessing interobserver variability was beyond the scope of this study. However, the scores of the two observers in this study were highly concordant for all assays, especially for CPS.

Thirdly, the patients in the cohort that were used in this study were treated with chemoradiotherapy, and not with immunotherapy. Therefore, this study is not a clinical validation study and no conclusions can be drawn on the predictive value of PD-L1 for immunotherapy and on the validity of the chosen cutoffs.

In conclusion, the results of this study do not support the hypothesis that the SP263 and the 22C3 standardized assays can be used interchangeably for determining PD-L1 expression in HNSCC. If these two different immunohistochemical assays are used in clinical decision making, it is critical that the diagnostic performance of the assays is highly comparable. As long as this cannot be confirmed, focus should be on the harmonization of the assays and caution must be taken when using PD-L1 expression to guide clinical practice.

Acknowledgements This study was supported by the Dutch Cancer Society (project number: A6C 7072).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Cohen EEW, Soulieres D, Le Tourneau C, Dinis J, Licitra L, Ahn MJ, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. Lancet. 2019;393:156–7.
2. Ferris RL, Blumenschein G Jr., Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med. 2016;375:1856–67.
3. Ferris RL, Blumenschein G Jr., Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. Oral Oncol. 2018;81:45–51.
4. Haddad R, Concha-Benavente F, Blumenschein G Jr, Fayette J, Guigay J, Colevas AD, et al. Nivolumab treatment beyond RECIST-defined progression in recurrent or metastatic squamous cell carcinoma of the head and neck in CheckMate 141: a subgroup analysis of a randomized phase 3 clinical trial. Cancer. 2019;125:3208–18.
5. Harrington KJ, Ferris RL, Blumenschein G Jr, Colevas AD, Fayette J, Licitra L, et al. Nivolumab versus standard, single-agent therapy of investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck (CheckMate 141): health-related quality-of-life results from a randomised, phase 3 trial. Lancet Oncol. 2017;18:1104–15.
6. Burtness B, Harrington KJ, Greil R, Soulieres D, Tahara M, de Castro G Jr, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. Lancet. 2019;394:1915–28.
7. Larkins E, Blumenthal GM, Yuan W, He K, Sridhara R, Subramaniam S, et al. FDA approval summary: Pembrolizumab for the treatment of recurrent or metastatic head and neck squamous cell carcinoma with disease progression on or after platinum-containing chemotherapy. Oncologist. 2017;22:873–8.
8. Soulieres D, Cohen E, Tourneau CL, Dinis J, Licitra L, Ahn MJ, et al. Updated survival results of the KEYNOTE-040 study of pembrolizumab vs standard-ofcare chemotherapy for recurrent or metastatic head and neck squamous cell carcinoma. Cancer Res. 2018;78:13.
9. Oliva M, Spreafico A, Taberna M, Alemany L, Coburn B, Mesia R, et al. Immune biomarkers of response to immune-checkpoint inhibitors in head and neck squamous cell carcinoma. Ann Oncol. 2019;30:57–67.
10. Sholl LM, Aisner DL, Allen TC, Beasley MB, Borczuk AC, Cagle PT, et al. Programmed Death Ligand-1 immunohistochemistry—A new challenge for pathologists: a perspective from members of the pulmonary pathology society. Arch Pathol Lab Med. 2016;140:341–4.

11. Ionescu DN, Downes MR, Christofides A, Tsao MS. Harmonization of PD-L1 testing in oncology: a Canadian pathology perspective. Curr Oncol. 2018;25:e209–e216.
12. Torlakovic E, Lim HJ, Adam J, Barnes P, Bigras G, Chan AWH, et al. "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod Pathol. 2020;33:4–17.
13. De Meulenaere A, Vermassen T, Creytens D, Aspeslagh S, Deron P, Duprez F, et al. Importance of choice of materials and methods in PD-L1 and TIL assessment in oropharyngeal squamous cell carcinoma. Histopathology. 2018;73:500–9.
14. Hodgson A, Slodkowska E, Jungbluth A, Liu SK, Vesprini D, Enepekides D, et al. PD-L1 immunohistochemistry assay concordance in urothelial carcinoma of the bladder and hypopharyngeal squamous cell carcinoma. Am J Surg Pathol. 2018;42:1059–66.
15. Ratcliffe MJ, Sharpe A, Rebelatto M, Scott M, Barker C, Scorer P, et al. A comparative study of PD-L1 diagnostic assays in squamous cell carcinoma of the head and neck (SCCHN). Ann Oncol. 2016;27:328–50.
16. van Kempen PM, van Bockel L, Braunius WW, Moelans CB, van Olst M, de Jong R, et al. HPV-positive oropharyngeal squamous cell carcinoma is associated with TIMP3 and CADM1 promoter hypermethylation. Cancer Med. 2014;3:1185–96.
17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
18. Munari E, Rossi G, Zamboni G, Lunardi G, Marconi M, Sommaggio M, et al. PD-L1 Assays 22C3 and SP263 are not interchangeable in non-small cell lung cancer when considering clinically relevant cutoffs: an interclone evaluation by differently trained pathologists. Am J Surg Pathol. 2018;42:1384–9.
19. Scott ML, Scorer P, Lawson N, Ratcliffe MJ, Barker C, Rebelatto M, et al. Assessment of heterogeneity of PD-L1 expression in NSCLC, HNSCC, and UC with Ventana SP263 assay. J Clin Oncol. 2017;35:e14502.