DR ALBINO ECCHER (Orcid ID : 0000-0002-9992-5550)

Article type : Review

Challenges Facing Pathologists Evaluating PD-L1 in Head & Neck Squamous Cell Carcinoma

Running title: IHC assays for PD-L1 in HNSCC

Ilaria Girolami MD[1], Liron Pantanowitz MD MHA[2], Massimo Barberis MD[3], Gaetano Paolino MD[4], Matteo Brunelli MD[4], Elena Vigliar MD[5], Enrico Munari MD[6], Swati Satturwar MD[7], Giancarlo Troncone MD[5], Albino Eccher MD[4]

[1]Division of Pathology, Central Hospital Bolzano, Bolzano, Italy

[2]Department of Pathology & Clinical Labs, University of Michigan, Ann Arbor, MI, USA

[3]Division of Pathology, IEO European Institute of Oncology, Milan, Italy

[4]Department of Pathology and Diagnostics, University and Hospital Trust of Verona, Verona, Italy

[5]Department of Public Health, University of Naples Federico II, Naples, Italy

[6]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

[7]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

Address correspondence and reprint requests to: Albino Eccher, Department of Pathology and Diagnostics, University and Hospital Trust of Verona, P.le Stefani n. 1; 37126, Verona, Italy. Phone: +390458122161, Fax: +390458122011, e-mail: albino.eccher@aovr.veneto.it

Word count 3025 words excluding abstract (150 words), references and table

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/JOP.13220

This article is protected by copyright. All rights reserved

Author Manuscript

Keywords: combined positive score; head and neck squamous cell carcinoma; immunohistochemical assay; programmed death-ligand 1; review

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Funding

No funding was received for this work.

Conflict of interest

The authors declare that they do not have any conflict of interest.

Ethics statement

No institutional review board approval was needed, as no ethical issue is raised by literature reviews.

Abstract

Programmed death-ligand 1 (PD-L1) expression with combined positive score (CPS) ≥1 is required for administration of checkpoint inhibitor therapy in recurrent/metastatic head and neck squamous cell carcinoma (HNSCC). The 22C3 pharmDx Dako immunohistochemical assay is the one approved as companion diagnostic for pembrolizumab, but many laboratories work on other platforms and/or with other clones, and studies exploring the potential interchangeability of assays have appeared. After review of the literature, it emerges that the concordance among assays ranges from fair to moderate, with a tendence of assay SP263 to yield a higher quota of positivity and of assay SP142 to stain better immune cells. Moreover, pathologists achieve very good concordance in assessing PD-L1 CPS, particularly with SP263. Differences in terms of platforms, procedures and study design still preclude a quantitative synthesis of evidence and clearly further work is needed to draw stronger conclusions on the interchangeability of PD-L1 assays in HNSCC.

INTRODUCTION

This article is protected by copyright. All rights reserved

Author Manuscript

Head and neck squamous cell carcinoma (HNSCC) affects approximately 880,000 new patients each year worldwide and represents a leading cause of mortality in some countries[1]. Despite combined therapy that includes surgical resection with radiotherapy and/or chemotherapy, the 5-year overall survival has improved only modestly over the past three decades and is only 50-65%[2,3]. A turning point in the therapy of HNSCC ensued with the introduction of immunotherapy targeting the programmed death-1 (PD-1)/programmed death-ligand-1 (PD-L1) axis[4–6]. The US Food and Drug Administration (FDA) and European Medicines Agency (EMA) approved of the PD-1 inhibitors pembrolizumab and nivolumab for recurrent and metastatic disease. The administration of these drugs showed improved survival with reduced toxicity[7–10]. Combination chemotherapy with other PD-1/PD-L1 inhibitors atezolizumab, avelumab, cemiplimab, and durvalumab are currently being tested in clinical trials[5], and interest is growing also for chemopreventive application[11,12]. However, the therapeutic benefit of single patients being treated with immune checkpoint inhibitors (ICI) is often not as impressive as expected.

PD-L1 expression in tumor and immune cells has been evaluated in clinical trials in relation to therapeutic response[13]. PD-L1 expression can be assessed by means of immunohistochemistry (IHC) with a tumor proportion score (TPS), which is the proportion of positive tumor cells, or with a combined positive score (CPS), which is defined by the ratio of total positive tumor and immune to the total number of viable tumor cells. These scoring systems are variably applied in clinical trials. In the KEYNOTE-040 study, patients with recurrent or metastatic HNSCC treated with pembrolizumab showed significantly improved survival when their tumor biopsies expressed PD-L1 with TPS ≥50%[7]. PD-L1 expression showed the best predictive performance when using a CPS with a cutoff of ≥20, and a lower but significant benefit in survival when CPS was ≥1 for first-line treatment in the KEYNOTE-048 study[5]. Other studies also consider a cutoff of ≥1% for both TPS and CPS as clinically relevant[9]. Recently, post-hoc analysis of the KEYNOTE-040 trial showed that CPS ≥ 50 is equivalent to TPS ≥ 50% for predicting objective response rate, overall survival, and progression-free survival in HNSCC patients[14]. Differences in prediction of response with the two scoring systems were predicted to be tied to inclusion of immune cells in the CPS, as around 50–60% of HNSCC tumor cells express PD-L1 when assessed with TPS, but this percentage increases to 85% when considering both tumor and surrounding immune cells, as is measured with the CPS[13,15]. Consequently, CPS appears to be more sensitive at lower cut-offs of positivity, supporting the importance of PD-L1-positive immune cells[14]. Finally, international agencies included a CPS of ≥1 as a selection criterion for first-line treatment of recurrent and metastatic HNSCC with

This article is protected by copyright. All rights reserved

Author Manuscript

pembrolizumab. However, similarly to what transpired in lung cancer, there are different PD-L1 IHC assays, using different PD-L1 antibodies (22C3, 28-8, SP263, SP142, E1L3N), on different IHC platforms.

Varying antibody clones and platforms have been approved for each available PD-1 and PD-L1 inhibitor, making comparison amongst immunotherapy trials difficult[16]. Two of these clones, 22C3 and 28-8, both run on the Dako immunohistochemistry platform, are the approved companion and complementary diagnostic test for pembrolizumab and nivolumab, respectively[17,18]. However, many institutions routinely work on other platforms (e.g. Roche), creating a problem as they are accordingly not then able to offer these originally approved assays. A laboratory developed test (LDT) is any test differing from the original regulatory approved commercial PD-L1 clone assay, no matter how small the difference is to one or more of its components/procedures [19]. Unfortunately, LDT development can be limited by an inability to standardize many of the assay components. As a result, LDTs are likely to be less robust than commercial tests and may thus introduce variability to results[19]. As occurred previously with lung cancer, harmonization studies among different assays and studies were necessary to compare their diagnostic performance. Herein we review the current field regarding the assessment of PD-L1 in HNSCC, compare these varied assays, and underscore challenges facing pathologists required to interpret PD-L1 expression.

METHODS

The review question was modeled on a Population, Intervention/Index, Comparator, Outcome (PICO) model. Population was represented by HNSCC only cases, Index test was any PD-L1 IHC clone on any platform, and Comparator term was considered any other clone or platform deployed to evaluate PD-L1 staining. We were interested in two outcomes: the concordance among different assays and secondarily the concordance among reading pathologists with any of the assays. A systematic search was carried out in Pubmed and Embase electronic database until January 10, 2021 with the combination of the key terms "PD-L1" and "HNSCC" defined with all their aliases. We did not insert an Outcome term in the search strategy to keep it as broad as possible. The complete search strategy is found in Supplementary material, Table S1. Inclusion criteria were the presence of any type of comparison among two or more assays for IHC PD-L1 expression in HNSCC or the comparison among pathologists assessing PD-L1 with one or more

This article is protected by copyright. All rights reserved

assays and the presence of reported concordance measures of any type (intraclass correlation coefficient (ICC), percent agreement, Cohen's kappa and variants) as outcomes of interest. No language or type of paper restrictions were applied. Studies not dealing with HNSCC (wrong Population), not dealing with PD-L1 assessment in IHC (wrong Index) or not presenting any type of comparison were excluded.

Quality of studies was assessed with a modified QUADAS2 tool tailored on our review question. We removed the question on case-control study design as it did not pertain to our review question, and we added the question on presence of detailed technical procedure with clone and platform. The modified set of items is found in Supplementary material Table S2.

RESULTS

Of 3379 items after removal of duplicates, after screening according to title and abstract 180 were assessed in full-text form and 19 were considered informative and included. The flow of article screening is found in Figure 1. The studies were represented by 10 full articles and 9 abstracts. The studies dealt with HNSCC cases from USA and Canada (n = 10, 53%), Europe (n = 7, 37%), Israel (n =1, 5%) and Latin America (n =1, 5%). The clone 22C3 was used in 14 (74%) studies, clone SP263 in 13 (68%), clone SP142 in 10 (53%), clone 28-8 in 5 (26%) and clone E1L3N in 2 (11%). Eight studies specifically evaluated PD-L1 expression with CPS, while the others evaluated separately tumor and immune cells with different cut-offs. The summary of the studies is found in Table 1, while full list of references is found in Supplementary material Appendix.

The summary of quality appraisal is reported in Supplementary Figure S1. The main item with risk of bias was the selection of cases, given that often it was not specified whether the cases were randomly or consecutively selected. Moreover, a great quota of studies was represented by abstracts with limited reporting, thus leading to a high number of studies with unclear risk of bias.

Given the high heterogeneity of studies it was not possible to perform a meta-analysis; hence, we provide an evidence-based review with qualitative discussion of most relevant studies, dividing the discussion according to the main outcomes: the concordance among assays and among pathologists.

Author Manuscript

This article is protected by copyright. All rights reserved

CONCORDANCE AMONG ASSAYS

Compared to lung cancer, evidence for direct comparison among assays in HNSCC is scarce. De Ruiter at al. compared the reference 22C3 assay with a 22C3 laboratory developed test (LDT) and the SP263 assay in 143 cases[20]. A smaller series published by Crosta et al. tested the performance of five different PD-L1 protocols with clones SP142, SP263 and 22C3 on various platforms against the 22C3 pharmDx on 15 cases/30 cores[21]. The aforementioned studies utilized a tissue microarray (TMA) and were comprised of cases from several sites in the head and neck region, with PD-L1 assessed centrally by trained pathologists. These studies both applied clinically relevant cut-offs of CPS = 1 and = 20 which emerged from clinical trials, while previous studies were based on TPS with various cut-offs, with or without distinction between staining of tumor or immune cells. Considering the quota of cases that are labeled as positive with CPS1-20 or higher than 20, both studies point towards an increased positivity rate when using the SP263 Ventana assay in comparison with the reference 22C3 pharmDx assay. The SP263 assay showed the greatest quota of positive cases compared with the other assays as reported by Crosta et al., whilst the 22C3 assay on the Omnis Dako platform showed the lowest. Crosta et al. also evaluated the performance of the various assays against the reference test by means of sensitivity and specificity for positive CPS ≥1 and found that the SP142 clone on the Ventana Benchmark Ultra platform performed best with a sensitivity of 92% and 100% specificity. The sensitivity and specificity of the other 22C3 assays run on other platforms ranged from 79-88% sensitivity and 80-83% specificity, while the SP263 assay showed 96% sensitivity and 50% specificity. This implies there is either a high quota of false positive results or that cases labelled as negative with the reference assay were incorrectly evaluated as positive with the SP263 assay. However, the negative predictive value of the best-performing SP142 was only 67%, meaning that one out of three negative cases with SP142 is not truly negative with 22C3 pharmDx. However, the Crosta et al. study has some limitations, such as the very limited number of cases evaluated, notable quota of non-evaluable cases was included, and concordance measures stronger than simple percent agreement such as ICC and kappa were not reported for pairwise comparisons. On the other hand, De Ruiter at al. reported precisely on the ICC and Cohen's kappa for their three comparisons and found that concordance between 22C3 pharmDx and SP263 is lower than moderate with ICC and in the range of fair (0.20-0.40) with Cohen's kappa both at cut-off 1 and 20 with no significant increase at the highest cut-off. The LDT of the study, a 22C3 clone on the Ventana Benchmark Ultra platform, showed an ICC of at least moderate in both the comparisons and concordance kappa from fair to

Author Manuscript

This article is protected by copyright. All rights reserved

substantial with high variability. De Ruiter et al. tested the concordance among the three assays, not only with the clinically relevant cut-offs of CPS1 and 20, but also at the cut-offs of 1% and 50% with TPS, providing an opportunity for comparison with other studies. The concordance of SP263 and 22C3 LDT with 22C3 pharmDx were all in the range of moderate (0.40-0.60), and interestingly it decreased at a higher cut-off for both the assays against the reference assay.

The impression that evaluation of CPS with SP263 yields a higher quota of positive (CPS≥1) cases has also been encountered in recent abstracts[22,23] and previous studies focused primarily on concordance among pathologists. Studies arising from a Canadian group of researchers involved a broad population of cancer cases among where 27 HNSCC mixed-site cases were stained with 22C3 pharmDx, SP142 and SP263 and clone E1L3N and evaluated with cut-off CPS≥1[24] or different percentages of positive immune or tumor cells[24–26]. These Canadian studies note that when a lower cut-off is applied (as represented by the intrinsic formulation of CPS, where positive immune cells are added to the numerator while the denominator comprises only tumor cells) a greater quota of cases is labelled as positive[24]. They also found a high concordance with Cohen's kappa in the range of substantial (0.60-0.80) for all pairwise comparisons and excellent ICC (higher than 0.90) for all clones with exception of SP142 when specifically considering concordance on tumor cell staining[25]. This is, however, in line with findings in other cancer sites where SP142 is known to better stain immune cells than tumor cells[27]. Studies from De Meulenaere et al., where the SP142 assay is used parallel to the reference 22C3 even if with no direct comparison, a tendency emerged towards lower rate of positivity in tumor cells with the SP142 [28,29].

Few studies and published abstracts have investigated the assay 28-8 run on the Dako platform, in comparison with other available assays[30–33]. Even though only overall percent agreement (OPA) was reported, these studies suggest a high degree of concordance among 28-8 and other assays, with OPA ranging from 80-100%[30–33], correlation coefficient higher than 0.9[32], and kappa of 0.82 which is considered almost perfect[31]. The 28-8 clone was also tested on platforms different from the Dako Autostainer Link 48 platform approved as a complementary diagnostic[34]. In this study, the platforms used were Dako Autostainer Link 48, Dako Omnis, Leica Bond-III, and Ventana Benchmark Ultra. The authors found an acceptable coefficient of variation in the percentages of positively stained target cells, with OPA ranging from 83-100% specifically for HNSCC cases between the various platforms and the reference pharmDx kit, and increasing concordance for higher cut-offs[34]. Studies such as the one reported by Koppel et al. shed light on common problems in harmonization and comparability studies, because of the multiplicity of platforms

This article is protected by copyright. All rights reserved

Author Manuscript

deployed and staining protocols applied, especially when multicenter comparison studies are attempted. This yields additional uncertainty, including the critical evaluation of published evidence, as no true direct comparison among results from studies is possible.

In summary, the main published findings to date support a variable degree of concordance from fair to substantial among the reference assay 22C3 pharmDx and two alternate Ventana assays, with the caveat that SP263 is likely to label more cases as positive due to stronger and more diffuse staining of tumor cells, while SP142 stains immune cells more so than tumor cells which in some cases can result in a lower CPS. Whilst limited evidence is available concerning the 28-8 assay, available findings show very good concordance with this assay and all others so far tested.

CONCORDANCE AMONG PATHOLOGISTS

Another important point when assessing PD-L1 expression, especially with the formulation of CPS, is concordance among pathologists. Indeed, CPS is more complex and perhaps less intuitive than TPS, as it requires specific counting of tumor and immune cells to calculate the score. Not surprisingly, training in this regard has been shown to be important[35]. However, when pathologists are trained, the reproducibility among them appears to be high in assessing CPS, with ICC≥0.70[21,29] or excellent ≥0.90[20] with all of the assays including 22C3 pharmDx, SP263 and SP142. Some studies suggest better concordance for pathologists with the SP263 assay when assessing both CPS or separately counting tumor and immune cells, with an almost perfect kappa of 0.836 and in the range of moderate to substantial for the clones SP142, 22C3 and E1L3N[24]. Similar results with very high concordance among scoring pathologists with SP263 assay have recently appeared in some abstracts[36,37]. Ease of use of the SP263 assay was previously recognized in validation studies. For example, Rebelatto et al. demonstrated the ease of use when the SP263 assay was validated in a mixed cohort of lung cancer and HNSCC cases[38]. The assay not only showed good analytical performance in terms of precision and robustness against pre-analytical factors, but also very high inter-observer and intra-observer concordance. The authors stated that the SP263 assay has analytical and reading properties in the same range of the two approved 22C3 and 28-8 assays, thus meeting the requirement criteria for assay use. Important to note, whilst the approved assays for HNSCC 22C3 and 28-8 use a clone that detects the epitopes of PD-L1 located in the extracellular domain, the Ventana assays and the clone E1L3N detect epitopes within the intracellular domain. It has been recognized that antibodies raised against the cytoplasmic domain

This article is protected by copyright. All rights reserved

Author Manuscript

of PD-L1 offer better visualization of membrane PD-L1 compared with those raised against the extracellular domain[39]. This may explain, at least partially, the difference in staining pattern and the preferential staining of tumor cells with SP263.

Finally, an important aid for pathologists to interpret PD-L1 expression could come from artificial intelligence (AI). As stated in a recent review by Inge et al., AI algorithms are already being used to investigate the role of PD-L1, providing considerable insight into its expression and heterogeneity within the tumor microenvironment[40]. Even though the few studies to date which have investigated AI algorithms for PD-L1 scoring dealt with cancer types other than HNSCC, the findings are interesting when bearing in mind that automated algorithms could be trained to recognize different types of cells (tumor vs immune; with macrophages to be excluded from other cell types) and provide a more objective quantification of CPS. In an AI study assessing CPS with image analysis in gastric cancer, the authors showed that application of their IHC membrane algorithm for PD-L1 evaluation was 85% concordant with manual scoring[41]. Furthermore, PD-L1 AI scores were comparable with manual scoring in predicting patient response to pembrolizumab[41]. The challenges encountered by the algorithms in recognizing and discriminating immune cells to be counted are the same as those encountered by pathologists. Hence, caution is warranted. It is foreseeable that, as transpired with using image analysis to assess breast biomarkers[42], the incorporation of clinical outcome as one of the benchmarks for algorithms could enable development of more clinically relevant tools.

CONCLUSION

The approval of ICI for HNSCC triggered by positive expression of PD-L1, and concurrent availability of several assays with different characteristics, prompted researchers to undertake comparative studies and explore interchangeability of these assays. While several such studies have been published in the last five years, only a small proportion dealt with HNSCC cases and incorporated varying cut-offs for positivity with or without evaluating immune and tumor cells. Only in the last year have studies appeared with a research design clearly focused on comparing of two or more assays with the reference 22C3 assay, and employing the required CPS scoring system. Emerging results currently suggest a moderate degree of concordance between the reference standard and widely used SP263 Ventana assay. Moreover, differences in staining patterns among these assays are observed, with SP263 demonstrating more strong staining of tumor cells and SP142 better

This article is protected by copyright. All rights reserved

Author Manuscript

staining of immune cells. However, heterogeneity in these varied studies (e.g. preanalytical factors, platforms used, detection systems applied, one vs multiple pathologist assessment) hamper critical and robust comparison of their results. Clearly, further work is needed to draw stronger conclusions on the interchangeability of PD-L1 assays in HNSCC.

Author Manuscript

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. Published online February 4, 2021:caac.21660. doi:10.3322/caac.21660

2. Pulte D, Brenner H. Changes in Survival in Head and Neck Cancers in the Late 20th and Early 21st Century: A Period Analysis. *Oncologist*. 2010;15(9):994-1001. doi:10.1634/theoncologist.2009-0289

3. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Prim*. 2020;6(1):92. doi:10.1038/s41572-020-00224-3

4. Cohen EEW, Soulières D, Le Tourneau C, 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(10167):156-167. doi:10.1016/S0140-6736(18)31999-8

5. Burtness B, Harrington KJ, Greil R, 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(10212):1915-1928. doi:10.1016/S0140-6736(19)32591-7

6. Canning M, Guo G, Yu M, et al. Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy. *Front Cell Dev Biol*. 2019;7. doi:10.3389/fcell.2019.00052

7. Ferris RL, Blumenschein GJ, Fayette J, 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*.

This article is protected by copyright. All rights reserved

2018;81:45-51. doi:10.1016/j.oraloncology.2018.04.008

8. Harrington KJ, Ferris RL, Blumenschein G, 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(8):1104-1115. doi:10.1016/S1470-2045(17)30421-7

9. Ferris RL, Blumenschein GJ, Fayette J, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med*. 2016;375(19):1856-1867. doi:10.1056/NEJMoa1602252

10. Ferris RL, Licitra L, Fayette J, et al. Nivolumab in Patients with Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck: Efficacy and Safety in CheckMate 141 by Prior Cetuximab Use. *Clin cancer Res an Off J Am Assoc Cancer Res*. 2019;25(17):5221-5230. doi:10.1158/1078-0432.CCR-18-3944

11. Ries J, Agaimy A, Wehrhan F, et al. Importance of the PD-1/PD-L1 Axis for Malignant Transformation and Risk Assessment of Oral Leukoplakia. *Biomedicines*. 2021;9(2):194. doi:10.3390/biomedicines9020194

12. Girolami I, Pantanowitz L, Munari E, et al. Prevalence of PD-L1 expression in head and neck squamous precancerous lesions: a systematic review and meta-analysis. *Head Neck*. Published online June 22, 2020:hed.26339. doi:10.1002/hed.26339

13. Cramer JD, Burtness B, Ferris RL. Immunotherapy for head and neck cancer: Recent advances and future directions. *Oral Oncol*. 2019;99:104460. doi:10.1016/j.oraloncology.2019.104460

14. Emancipator K, Huang L, Aurora-Garg D, et al. Comparing programmed death ligand 1 scores for predicting pembrolizumab efficacy in head and neck cancer. *Mod Pathol an Off J United States Can Acad Pathol Inc*. Published online November 2020. doi:10.1038/s41379-020-00710-9

15. Concha-Benavente F, Srivastava RM, Trivedi S, et al. Identification of the Cell-Intrinsic and -Extrinsic Pathways Downstream of EGFR and IFNγ That Induce PD-L1 Expression in Head and Neck Cancer. *Cancer Res*. 2016;76(5):1031-1043. doi:10.1158/0008-5472.CAN-15-2001

16. Ancevski Hunter K, Socinski MA, Villaruz LC. PD-L1 Testing in Guiding Patient Selection for

This article is protected by copyright. All rights reserved

Author Manuscript

PD-1/PD-L1 Inhibitor Therapy in Lung Cancer. *Mol Diagn Ther*. 2018;22(1):1-10. doi:10.1007/s40291-017-0308-6

17. PD-L1 IHC 22C3 pharmDx Interpretation Manual – Head and Neck Squamous Cell Carcinoma (HNSCC). https://www.agilent.com/cs/library/usermanuals/public/29314_22c3_pharmDx_hnscc_interpretation_manual_us.pdf

18. PD-L1 IHC 28-8 pharmDx Interpretation Manual – Head and Neck Squamous Cell Carcinoma (HNSCC). https://www.agilent.com/cs/library/usermanuals/public/29186_pd-l1-ihc-28-8-interpretation-manual-scchn.pdf

19. Ionescu DN, Downes MR, Christofides A, Tsao MS. Harmonization of PD-L1 testing in oncology: a Canadian pathology perspective. *Curr Oncol*. 2018;25(3):e209-e216. doi:10.3747/co.25.4031

20. de Ruiter EJ, Mulder FJ, Koomen BM, et al. Comparison of three PD-L1 immunohistochemical assays in head and neck squamous cell carcinoma (HNSCC). *Mod Pathol an Off J United States Can Acad Pathol Inc*. Published online August 2020. doi:10.1038/s41379-020-0644-7

21. Crosta S, Boldorini R, Bono F, et al. PD-L1 Testing and Squamous Cell Carcinoma of the Head and Neck: A Multicenter Study on the Diagnostic Reproducibility of Different Protocols. *Cancers (Basel)*. 2021;13(2):292. doi:10.3390/cancers13020292

22. Ramos M, Baldion A, Suarez D, et al. Immunohistochemistry analysis of PD-L1 expression in head and neck cancer. *Virchows Arch*. 2019;475(S1):1-436. doi:10.1007/s00428-019-02631-8

23. Al-Masri H, Ratcliffe M, Sharpe A, et al. Concordance of tumour and immune cell staining with Ventana SP142, Ventana SP263, Dako 22C3 and Dako 28–8 PD-L1 tests across different cancer types. *Virchows Arch*. 2017;471(S1):1-352. doi:10.1007/s00428-017-2205-0

24. Downes MR, Slodkowska E, Katabi N, Jungbluth AA, Xu B. Inter- and intraobserver agreement of programmed death ligand 1 scoring in head and neck squamous cell carcinoma, urothelial carcinoma and breast carcinoma. *Histopathology*. 2020;76(2):191-200. doi:10.1111/his.13946

This article is protected by copyright. All rights reserved

Author Manuscript

25. Hodgson A, Slodkowska E, Jungbluth A, et al. PD-L1 Immunohistochemistry Assay Concordance in Urothelial Carcinoma of the Bladder and Hypopharyngeal Squamous Cell Carcinoma. *Am J Surg Pathol*. 2018;42(8):1059-1066. doi:10.1097/PAS.0000000000001084

26. Wang C, Hahn E, Slodkowska E, et al. Reproducibility of PD-L1 immunohistochemistry interpretation across various types of genitourinary and head/neck carcinomas, antibody clones, and tissue types. *Hum Pathol*. 2018;82:131-139. doi:10.1016/j.humpath.2018.07.024

27. 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 Off Publ Int Assoc Study Lung Cancer*. 2017;12(2):208-222. doi:10.1016/j.jtho.2016.11.2228

28. De Meulenaere A, Vermassen T, Aspeslagh S, et al. Tumor PD-L1 status and CD8(+) tumor-infiltrating T cells: markers of improved prognosis in oropharyngeal cancer. *Oncotarget*. 2017;8(46):80443-80452. doi:10.18632/oncotarget.19045

29. De Meulenaere A, Vermassen T, Creytens D, et al. Importance of choice of materials and methods in PD-L1 and TIL assessment in oropharyngeal squamous cell carcinoma. *Histopathology*. 2018;73(3):500-509. doi:10.1111/his.13650

30. Gatalica Z, Vanderwalde AM, Rose I, et al. Distribution of PD-L1 expression in diverse cancer types: Experience with over 10,000 cases. *J Clin Oncol*. 2016;34(15_suppl):11548-11548. doi:10.1200/JCO.2016.34.15_suppl.11548

31. Krigsfeld GS, Prince EA, Pratt J, et al. Analysis of real-world PD-L1 IHC 28-8 and 22C3 pharmDx assay utilisation, turnaround times and analytical concordance across multiple tumour types. *J Clin Pathol*. 2020;73(10):656-664. doi:10.1136/jclinpath-2020-206466

32. Ratcliffe M, Sharpe A, Rebelatto M, et al. A comparative study of PD-L1 diagnostic assays in squamous cell carcinoma of the head and neck (SCCHN). *Ann Oncol*. 2016;27(6):vi328–vi350. doi:10.1093/annonc/mdw376.7

33. Scott M, Wildsmith S, Ratcliffe M, et al. Comparison of patient populations identified by different PD-L1 assays in head and neck squamous cell carcinoma (HNSCC). *Ann Oncol*. 2018;29(8):375.

Author Manuscript

This article is protected by copyright. All rights reserved

34. Koppel C, Schwellenbach H, Zielinski D, et al. Optimization and validation of PD-L1 immunohistochemistry staining protocols using the antibody clone 28-8 on different staining platforms. *Mod Pathol an Off J United States Can Acad Pathol Inc*. 2018;31(11):1630-1644. doi:10.1038/s41379-018-0071-1

35. Eccher A, Fontanini G, Fusco N, et al. Digital slides as an effective tool for programmed death ligand 1 combined positive score assessment and training: Lessons learned from the "Programmed death ligand 1 key learning program in Head-and-Neck squamous cell carcinoma." *J Pathol Inform*. 2021;12(1):1. doi:10.4103/jpi.jpi_63_20

36. Jamshidi P, Kaur A, Liu L, Sullivan M, Watkin W, Paintal A. Inter-Rater Reliability (IRR) in a Consensus PD-L1 Immunohistochemistry (IHC) Service in an Academic Multi-Hospital Health System. *Mod Pathol*. 2020;33(suppl 2):1863.

37. Nielsen A, Manriquez G, Hayden D, et al. Precision and Repeatability of the VENTANA PD-L1 (SP263) Assay Across Six Different Tumor Types. *Mod Pathol*. 2020;33(suppl 2):830.

38. Rebelatto MC, Midha A, Mistry A, et al. Development of a programmed cell death ligand-1 immunohistochemical assay validated for analysis of non-small cell lung cancer and head and neck squamous cell carcinoma. *Diagn Pathol*. 2016;11(1):95. doi:10.1186/s13000-016-0545-8

39. Mahoney KM, Sun H, Liao X, et al. PD-L1 Antibodies to Its Cytoplasmic Domain Most Clearly Delineate Cell Membranes in Immunohistochemical Staining of Tumor Cells. *Cancer Immunol Res*. 2015;3(12):1308-1315. doi:10.1158/2326-6066.CIR-15-0116

40. Inge LJ, Dennis E. Development and applications of computer image analysis algorithms for scoring of PD-L1 immunohistochemistry. *Immuno-Oncology Technol*. 2020;6:2-8. doi:10.1016/j.iotech.2020.04.001

41. Kim H-N, Jang J, Heo YJ, et al. PD-L1 expression in gastric cancer determined by digital image analyses: pitfalls and correlation with pathologist interpretation. *Virchows Arch*. 2020;476(2):243-250. doi:10.1007/s00428-019-02653-2

42. Barnes M, Srinivas C, Bai I, et al. Whole tumor section quantitative image analysis maximizes between-pathologists' reproducibility for clinical immunohistochemistry-based biomarkers. *Lab Investig*. 2017;97(12):1508-1515. doi:10.1038/labinvest.2017.82

This article is protected by copyright. All rights reserved

Author Manuscript

Figure legends.

Figure 1. Flow of article screening according to PRISMA.

Author Manuscript

This article is protected by copyright. All rights reserved

Table 1. Summary of retrieved studies dealing with comparison among assays in HNSCC

Author, year (Country)	N cases	Sites	Scoring system	Clone	Main findings	Main limitations/Notes
Al-Masri, 2017 (USA)	200-500	HNSCC NOS	TC and IC, NOS	• SP142 Ventana • SP263, Ventana • 22C3, DAKO • 28-8, DAKO	• Strong intrinsic agreement in TC and IC PD-L1 expression between the SP263 and 22C3 and 28–8 assays. • IC scoring more variable than TCs • SP142 assay with lower agreement versus the other assays in both NSCLC and HNSCC	• Abstract only, no quantitative measures reported • Also NSCLC cases
Crosta, 2021 (Italy)	15 (30 cores)	OC, P, L	CPS with 1 and 20 cut-offs	• 22C3 pharmDx Dako Autostainer as gold standard • 22C3 Ventana Benchmark • SP142 Ventana Benchmark • 22C3 Leica Bond • SP263 Ventana Benchmark • 22C3 Dako Omnis	• Best performance with the SP142 clone on the Ventana Benchmark Ultra platform (92% and 100% sensitivity and specificity) • Sensitivity (88-96%) and specificity (50-100%) varying among protocols of staining • Higher proportion of positive cases with SP263 (sensitivity 96%, specificity 50%) • ICC 0.774 for inter-observer reliability and fair to moderate agreement for the cut-offs	• Few cases and some selection bias • Great quota of unevaluable cases • Concordance measures are not reported for pairwise comparisons

Author Manuscript

This article is protected by copyright. All rights reserved

De Meulenaere, 2017 (Belgium)	99	OP	Positive if ≥5% TC or IC	• 22C3 Agilent-Dako 1:100 with Ultraview kit Ventana • SP142 Roche with Optiview kit Ventana	• More TC positive with 22C3 than SP142 (34% vs 23%) • Low levels of positivity in IC (2-3%) • Positivity in TC strongly linked with positivity in immune environment with both clones • OS associated with PD-L1 expression with SP142 clone only	• No use of scoring systems used in trials (TPS or CPS) • No direct comparison among clones
De Meulenaere, 2018 (Belgium)	99	OP	Cut-offs 1%, 5%, 10%	• 22C3 Agilent-Dako 1:100 with Ultraview kit Ventana • SP142 Roche with Optiview kit Ventana	• Moderate agreement (κ 0.511) among clones and OPA 75-83.3% at various cut-offs • Substantial agreement among four scoring pathologists (ICC >0.70) for both clones • OPA among pathologists at various cut-offs 66-81% for SP142 and 65%-74% for 22C3 • Strong correlation between manual and digital assessment of PD-L1 for both clones	No special limitations; the study explores also the agreement between different type of specimens (resection vs biopsy, primary vs node metastasis), finding that biopsy underscores PD-L1 expression with both clones
De Ruiter, 2020 (Netherlands)	147	OP, HP, L	TPS and CPS	• 22C3 pharmDx assay on the Dako Link 48	• ICC always lower than moderate (<0.50) and κ in the range of poor or fair for the relevant cut-offs among	No special limitations; the study reports also a good concordance among two

This article is protected by copyright. All rights reserved

Author Manuscript

Author Manuscript

				• SP263 assay on the Ventana Benchmark Ultra • 22C3 as an LDT on Ventana Benchmark Ultra	SP263 and 22C3 pharmDx for both TPS and CPS • For CPS, concordance of 22C3 LDT with the SP263 assay and with the 22C3 pharmDx could be defined as moderate to poor • For TPS, 22C3 LDT more concordant with the SP263 assay than with the 22C3 pharmDx assay • Concordance is slightly increasing at higher cut-offs with CPS • SP263 tends to stain more cases than 22C3	scoring pathologists, a higher agreement between TMA and whole sections for TPS than CPS and significant intratumor heterogeneity
Downes, 2020 (Canada)	27	HNSCC NOS	CPS≥1, IC≥1%, TC≥25%	• SP142 on Ventana Benchmark Ultra • SP263 on Ventana Benchmark Ultra • 22C3 pharmDX on a Dako Autostainer Link 48	• When a lower threshold is used (CPS with 22C3) a higher proportion of cases is positive • Interobserver OPA 78-96% and κ 0.517-0.836 higher with SP263, lower with SP142) • Intraobserver agreement almost perfect with all assays	• No direct comparison among assays • All the cut-offs are applied to E1L3N • Similar findings also in the cohorts of breast and urothelial cancer

This article is protected by copyright. All rights reserved

Author Manuscript

				• E1L3N on Leica Bond-III Autostainer		
Frederick, 2020 (USA)	NS	HNSCC NOS	CPS 1, 10, 20	22C3 pharmDx	• Survey of studies on intraobserver and interobserver concordance among 42 pathologists after training in assess CPS • 73% and 93% of pathologist met all the endpoint for intra- and interobserver reproducibility	• Abstract only • Several types of cancer other than HN
Gatalica, 2016 (USA)	NA	HNSCC NOS	5% cut-of	• SP142, Spring Biosciences • SP263, Ventana • 22C3, DAKO • 28-8, DAKO	Reported 90% concordance among the 4 assays	• Abstract only • Several cancer types other than HNSCC
Hodgson, 2018 (Canada)	27	HP	Positive if ≥25% TC with SP263, or if ≥50% TC or 10% IC with SP142, or	• SP142 on the Ventana Benchmark Ultra • SP263 on the Ventana Benchmark Ultra • E1L3N on Ventana	• Substantial agreement between any of the two antibody clones compared with kappa ranging 0.639-0.791 • High ICC among assays for TC (≥0.90), lower for IC (ICC 0.51-0.86) • SP142 scoring highest IC, SP263 scoring highest TC	• Results for both HN SCC and urothelial cancer together • TMA only

This article is protected by copyright. All rights reserved

			if ≥50% TC with 22C3	Benchmark Ultra • All three with the Optiview DAB IHC detection kit for all the clones, followed by the Optiview amplification kit for SP142 only • 22C3 with the DAKO EnVision FLEX system on a DAKO Autostainer Link 48 system		
Jamshidi, 2020 (USA)	22	HNSCC NOS	CPS 1 and 20	SP263	• 95% concordance and κ=0.78 at CPS1 • 86% concordance and κ=0.81 at CPS20 • Four pathologists scoring	Abstract only
Koppel, 2018 (Germany)	30	HNSCC and ADK	TC cut-offs 1%, 5%, 10%, 50%	Same clone (28-8) with different platforms: • pharmDx kit Autostainer Link	• 100% inter- and intra-assay repeatability scoring/classification, with acceptable coefficient of variation in percentages of positively	• Also NSCLC and melanoma cases • Not explored interobserver

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

				48 • Autostainer Link 48 • Dako Omnis • Leica Bond-III and Bond Polymer Refine Detection • Ventana Benchmark ULTRA with OptiView DAB IHC Detection Kit	stained target cells • OPA ranging 83-100% for HNSCC cases between the various platforms and the reference pharmDx kit • Higher concordance for higher cut-offs	concordance • TMA study • TC only
Krigsfeld, 2020 (USA)	305	HNSCC NOS	Cut-offs 1%, 5%, 10%, 25%, 50%	• 28-8 pharmDx Dako • 22C3 pharmDx Dako • SP142 Ventana	Agreement ranging 81-100% between 28-8 and 22C3 with κ=0.82 at cut-off 1%	Also NSCLC, melanoma and urothelial cancer with similar agreement rates and correlation
Nielsen, 2020 (USA)	NA	HNSCC NOS	NA	SP263 on three platforms	98% OPA inter- and intraobserver 99% OPA among platforms	Abstract only; type of cases and specimens not declared
Ramos, 2019 (Colombia)	38	HNSCC NOS	TPS and CPS≥1	• Ventana SP263 • Dako 22C3	Ventana's antibody showed a more intense staining, facilitating the overall assessment of PD-L1 expression	Abstract only, no reported quantitative measures
Ratcliffe, 2016	108	HNSCC	1%, 10%,	• SP263 Ventana	• Correlation coefficient of ≥0.9 for	Abstract only

This article is protected by copyright. All rights reserved

Author Manuscript

(UK)		NOS	25%	• 28-8 Dako • 22C3 Dako	each pairwise comparison • OPA >90% among the three assays across multiple clinically relevant cut points	
Rebelatto, 2016 (USA)	100	HNSCC NOS	Positive if ≥25% TC	• SP263 Ventana Benchmark Ultra platform and with the Optiview DAB Detection Kit	• OPA 90.8% between three pathologists • Intra-observer OPA 94.3 % • Good analytical properties of the assay in terms of precision, in the range of the 22C3 and 28-8 assays	Also NSCLC cases, where the cut-off of 25% positive TC with SP263 can discriminate responders to durvalumab
Scott, 2018 (UK)	486	HNSCC NOS	TC, IC and CPS 1 and 10	• SP263 Ventana • SP142 Ventana • 22C3 pharmDx • 28-8 pharmDx	• Good analytical correlation for TC staining for SP263, 22C3 and 28-8 assays • SP263 appearing more sensitive, with higher proportion of positive cases • OPA between SP263 and other assays at CPS≥1 ranging 69-83%	Abstract only
Vainer, 2019 (Israel)	11	HNSCC NOS	CPS≥1	• 22C3 antibody LDT for the BenchMark XT • 22C3 pharmDx	• ICC 0.83 among the two assays	Abstract only
Wang, 2018 (USA)	49	HNSCC HP	Positive if ≥25% TC	• SP263 • SP142	• Moderate interobserver agreement for both SP263 (κ = 0.469) and SP142	• No CPS, cut-offs for urothelial cancer

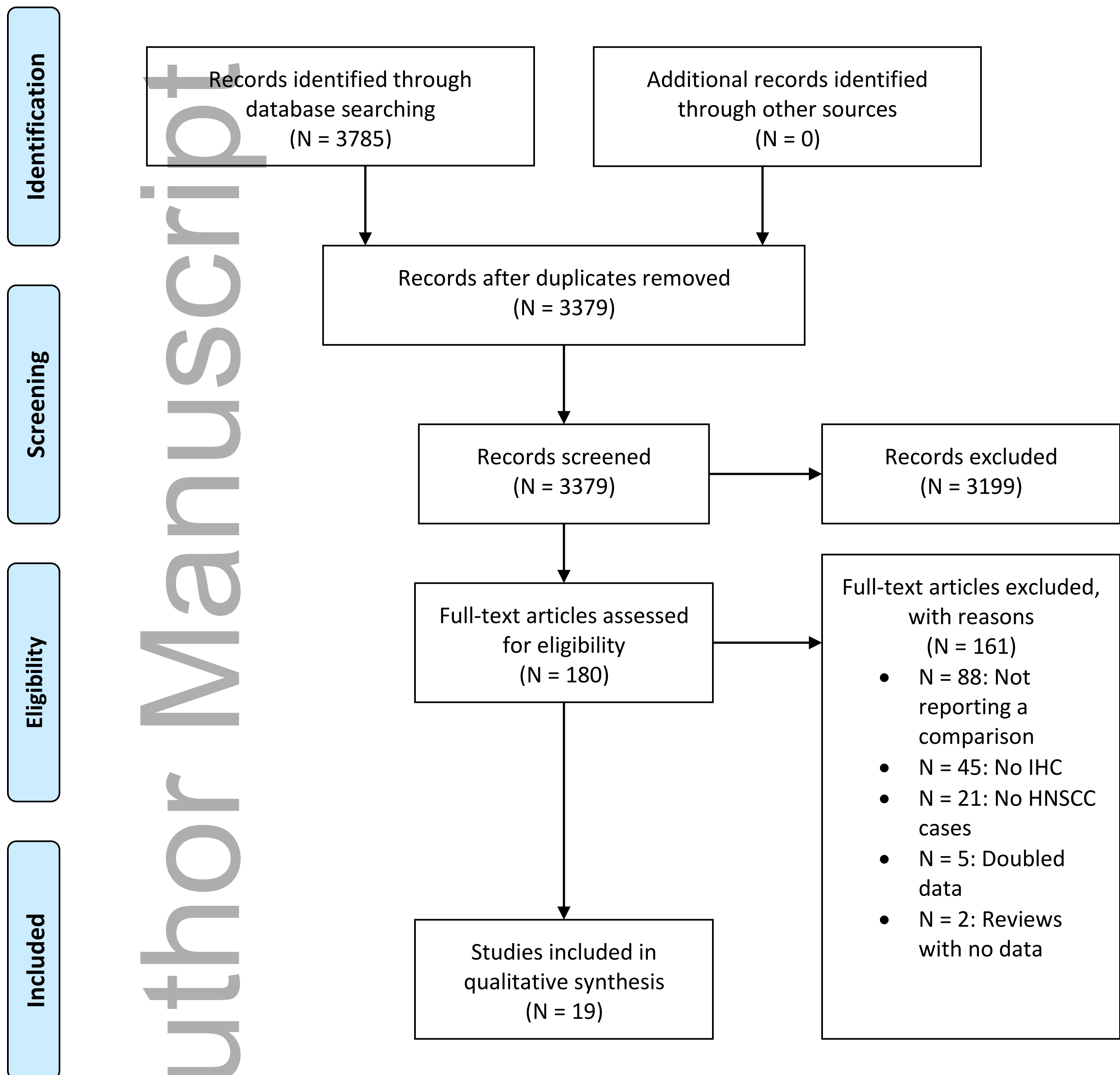
This article is protected by copyright. All rights reserved

	and SGC	or IC with SP263, if ≥5% IC for SP142	• Both on Benchmark Ultra platform and with the Optiview DAB Detection Kit followed by the Optiview Amplification Kit for SP142 only	clones (κ = 0.591) • OPA 80% for SP263 and 88% for SP142	applied to HNSCC • No comparison among clones • Additional finding of substantial agreement between TMA and WS in (κ= 0.667)

ADK, adenocarcinoma; CPS, combined proportion score; HNSCC, head&neck squamous cell carcinoma; HP, hypopharyngeal; IC, immune cells; L, laryngeal; LDT, laboratory developed test; OC, oral cavity; OP, oropharynx; OPA, overall percent agreement; NOS, not otherwise specified; NSCLC, non-small cell lung cancer; SGC, salivary gland cancer; TC, tumor cells; TMA, tissue micro array; TPS, tumor proportion score; WS, whole section

Author Manuscript

This article is protected by copyright. All rights reserved



This article is protected by copyright. All rights reserved