Interassay and interobserver comparability study of four programmed death-ligand 1 (PD-L1) immunohistochemistry assays in triple-negative breast cancer

Aurelia Noske [a, *], Daniel-Christoph Wagner [b], Kristina Schwamborn [a], Sebastian Foersch [b], Katja Steiger [a], Marion Kiechle [c], Dirk Oettler [d], Siranush Karapetyan [e], Alexander Hapfelmeier [e, f], Wilfried Roth [b], Wilko Weichert [a]

[a] Institute of Pathology, School of Medicine, Technical University of Munich, Munich, Germany
[b] Institute of Pathology, University Medical Centre, Johannes Gutenberg University Mainz, Mainz, Germany
[c] Department of Gynaecology and Obstetrics, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
[d] MSD Sharp & Dohme GmbH, Haar, Germany
[e] Institute of General Practice and Health Services Research, School of Medicine, Technical University of Munich, Munich, Germany
[f] Institute of Medical Informatics, Statistics and Epidemiology, School of Medicine, Technical University of Munich, Munich, Germany

ARTICLE INFO

ABSTRACT

Different immunohistochemical programmed death-ligand 1 (PD-L1) assays and scorings have been reported to yield variable results in triple-negative breast cancer (TNBC). We compared the analytical concordance and reproducibility of four clinically relevant PD-L1 assays assessing immune cell (IC) score, tumor proportion score (TPS), and combined positive score (CPS) in TNBC. Primary TNBC resection specimens (n = 104) were stained for PD-L1 using VENTANA SP142, VENTANA SP263, DAKO 22C3, and DAKO 28–8. PD-L1 expression was scored according to guidelines on virtual whole slide images by four trained readers.

The mean PD-L1 positivity at IC-score ≥1% and CPS ≥1 ranged between 53% and 75% with the highest positivity for SP263 and comparable levels for 22C3, 28–8, and SP142. Inter-assay agreement was good between 28–8 and 22C3 across all scores and cut-offs (kappa 0.68–0.74) and for both assays with SP142 at IC-score ≥1% and CPS ≥1 (kappa 0.61–0.67). The agreement between SP263 and other assays was substantially lower for all scores. Inter-reader agreement for each assay was good to excellent for IC-score ≥1% (kappa 0.73–0.78) and CPS ≥1 (kappa 0.68–0.74), fair to good for CPS ≥10 (kappa 0.52–0.67) and TPS ≥1% (kappa 0.53–0.72). The percentage of overlapping cases in the positive/negative category was >90% between IC-score ≥1% and CPS ≥1 but below when comparing IC-score ≥1% with CPS ≥10. We demonstrate an overall good inter-reader agreement for all PD-L1 assays in TNBC along with assay specific differences in positivity and concordances, which may aid to select the right test strategy in routine diagnostics.

1. Introduction

Triple-negative breast cancer (TNBC) lacks oestrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER2) overexpression/amplification. This aggressive breast cancer subtype has a high risk of disease progression. Treatment options are limited, but relevant progress has been made with the approval of novel targeted therapies, such as poly-ADP ribose polymerase (PARP) inhibitors for patients with BRCA 1/2 germline mutations, the antibody-drug conjugate Sacituzumab

govitecan, and immune checkpoint blockade [1]. TNBC is a heterogeneous disease, and a subset of these tumors is characterised by specific interactions with microenvironmental factors like immune cells [2]. A prominent clinically relevant example is that a subset of TNBC show varying levels of programmed death-ligand 1 (PD-L1)-positive immune and tumor cells which influence response to immune checkpoint inhibitors (ICI).

Atezolizumab plus nab-paclitaxel is a first-line treatment option approved by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for patients with inoperable or metastatic TNBC who have a PD-L1 positive immune cell (IC) score $\geq$1% based on data from the IMpassion130 trial [3]. The benefit of atezolizumab was demonstrated especially in the PD-L1 IC (SP142 assay) positive population [3,4].

Pembrolizumab in combination with chemotherapy is approved by the FDA (and EMA's acceptance is expected soon) for the treatment of advanced or metastatic TNBC with PD-L1 CPS $\geq$10 (22C3 assay) status. According to data from the Keynote-355 trial, these patients have a significant improvement in PFS when pembrolizumab is added to chemotherapy as compared to the placebo group [5].

In early TNBC, neoadjuvant atezolizumab plus standard chemotherapy revealed a higher pathological complete response (pCR) in both the intention-to-treat (ITT) and the PD-L1 positive population regardless of the PD-L1 IC (SP142) status according to the Impassion031 trial [6]. Similarly, neoadjuvant treatment with pembrolizumab plus chemotherapy versus chemotherapy alone showed a higher pCR in Keynote-522 [7], here the benefit was independent from the PD-L1 CPS (22C3) status. According to these clinical trials, the predictive value of PD-L1 might vary between early and advanced TNBC but further studies to address this topic are needed.

Based on the study data outlined above, currently, the assessment of the PD-L1 status is mandatory for the identification of TNBC patients eligible for ICI therapies. In contrast to the FDA, the EMA does not directly link a drug to one specific assay and staining platform for PD-L1 testing. European and national guidelines for the treatment of breast cancer recommend PD-L1 testing [8,9]. However, the analysis of PD-L1 is challenging since several antibodies, platforms and scoring criteria are available. It is important that pathologists ensure the use of a valid test and have experience in the interpretation of staining results. To be able to decide which assay can be safely used, antibody and staining platform comparability studies are mandatory. As the equivalence of PD-L1 assays and scorings is still a matter of debate, we aimed to investigate the analytical comparability and reproducibility of four clinically recommended PD-L1 IHC assays evaluated by four trained readers using different scoring methods in a large cohort of resected TNBC.

2. Materials and methods

2.1. Study population

This study was designed to assess inter-assay and inter-reader comparability of PD-L1-status across four PD-L1 IHC assays in primary TNBC. Archival, formalin-fixed, paraffin-embedded, resection specimens (n = 104) from the Institute of Pathology, Technical University of Munich (TUM), Germany were enrolled as described previously [10]. All samples were negative for hormone receptors and HER2 according to ASCO/CAP guidelines [11,12]. The median age at time of diagnosis was 53 years (range, 27–89). Pathological tumor stage was available in 85 cases. The most frequent tumor stage was pT2 in 44.7% followed by pT1c in 36.5%. Nodal stage was known in 57 cases and a nodal negative status was present in 56.1%. Clinical outcome data were not available. Tissue processing and use

was coordinated within the framework of the Klinikum rechts der Isar/TUM tissue biobank (subject to strict legal and ethical regulations). The investigation complied with the current laws of the country in which they were performed.

2.2. PD-L1 IHC assays

Immunohistochemistry was conducted with four PD-L1 antibodies on two different staining platforms. The VENTANA SP142 (Roche Diagnostics, Mannheim, Germany) and the VENTANA SP263 assay (Roche Diagnostics) were used on the VENTANA Benchmark Ultra platform at TUM. PD-L1 IHC 22C3 pharmDx (Agilent Technologies, Waldbronn, Germany) and PD-L1 IHC 28-8 pharmDx (Agilent Technologies) assays were run on a DAKO Autostainer Link 48 at the Institute of Pathology, University Medical Centre Mainz (Germany). All assays are referred to hereafter by the clone of the antibody used.

2.3. Evaluation of PD-L1 staining and scoring

All PD-L1 stained slides and corresponding haematoxylin and eosin (H&E) stains were digitized (Leica Aperio AT2, TUM), stored into a database (Aperio eSlide Manager) and virtually assessed by four pathologists (AN, DCW, KS, SF). Access to the slides was randomized and blinded for patient and assay information on the digital platform. Each pathologist investigated in total 396 slides. PD-L1 expression was evaluated in immune cells (IC) and tumor cells (TC) separately for each assay. Staining was scored for IC-positivity as the percentage of invasive tumor area covered by stained immune cells (defined as staining in granulocytes, lymphocytes, macrophages, and dendritic cells of any intensity) [13]. The tumor proportion score (TPS) was evaluated according to the percentage of stained viable TC in the tumor area showing partial or complete membranous PD-L1 staining of any intensity. The combined positive score (CPS) was calculated by summing the number of PD-L1 stained cells (TC, IC) and dividing the sum by the total number of viable tumor cells, multiplied by 100 [14].

2.4. Statistics

Fleiss' Kappa for multiple readers and the intraclass correlation ICC(2, k) was used to quantify agreement of categorical and continuous measurements, respectively. Kappa values and ICC coefficients were interpreted according to the guideline of Cicchetti [15]. The probabilities of PD-L1-positivity for IC-score, TPS or CPS across assays and corresponding 95% confidence intervals were estimated by linear mixed effects regression models with a fixed factor variable for assays and random intercepts for observers and patients. All analyses have been conducted in R 4.0.3 (The R Foundation for Statistical Computing, Vienna, Austria).

3. Results

3.1. Prevalence of PD-L1 expression

High quality PD-L1 staining with all assays was available in 99 TNBC samples. Of these, due to the lack of invasive tumor, five cases had to be excluded from the final evaluation. The adjusted mean percentages of PD-L1 positivity for each assay and for the three scores across all readers are summarized in Table 1/Fig. 1. The overall IC-score ($\geq$1%) positivity rates were similar for SP142, 22C3, and 28-8 but higher for SP263. TPS ($\geq$1%) positivity rates were similar for 22C3 and 28–8, but higher for SP263 and lower for SP142. PD-L1-positivity rates according to CPS $\geq$1 and CPS $\geq$10 were similar for SP142, 22C3, and 28-8 but higher for SP263. In

Table 1
Adjusted mean percentages of PD-L1 positivity for IC-score, TPS, and CPS across all samples for each assay.

Assay	IC-score ≥ 1% (95% CI)	TPS ≥1% (95% CI)	CPS ≥1 (95% CI)	CPS ≥10 (95% CI)
SP142	53% (43–62%)	9% (0.4–18%)	56% (44–67%)	17% (6–27%)
SP263	74% (64–83%)	43% (33–52%)	75% (63–86%)	36% (26–47%)
22C3	53% (43–62%)	25% (15–34%)	53% (41–64%)	17% (7–28%)
28–8	59% (49–68%)	29% (19–38%)	59% (48–70%)	24% (13–34%)
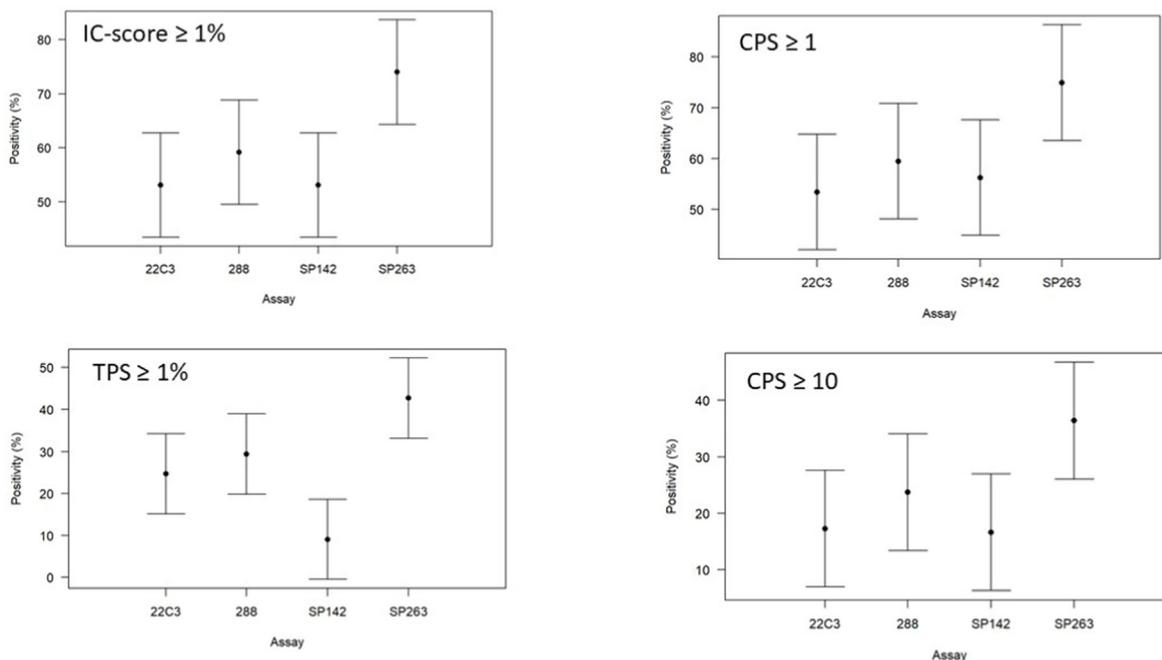
Fig. 1. Adjusted mean percentages of PD-L1-positivity for each assay and different scoring methods across the four readers are depicted with CI 95%.

Fig. 2 we show ungrouped PD-L1 positivity rates for IC-score and CPS across all readers for each assay and sample.

3.2. Inter-assay PD-L1-positivity agreement

To test the inter-assay variability, we determined ICCs for assay pairs at each score averaged for all readers as summarized in Table 2. Of all assay combinations, the comparison of 28–8 vs. 22C3 showed good to excellent agreement for each score, while 28–8 vs. SP142 and 22C3 vs. SP142 demonstrated good agreement only for the IC-Score. Agreement of SP263 with any of the other assays was poor to fair in almost all scenarios, due to generally higher PD-L1 staining scores seen for this antibody. Next, we determined Fleiss'

kappa to assess the reliability of multiple measurements when fixed cut-offs were applied. We compared pair-wise assay combinations at the specific cut-offs averaged for all readers as shown in Table 3. Of all assay pairs, 28–8 vs. 22C3 showed a good agreement for each score. Agreement between SP142 and either 22C3 or 28–8 was also good for IC and CPS 1 but not for TPS and CPS 10. Again, agreement of SP263 with any of the other assays was poor to fair.

3.3. Inter-reader PD-L1-positivity agreement

Evaluation of ICCs to quantify inter-reader variability for each assay and IC-score were 0.489–0.793 (fair to excellent inter-reader agreement) and highest for SP142. ICCs for CPS were 0.653–0.794
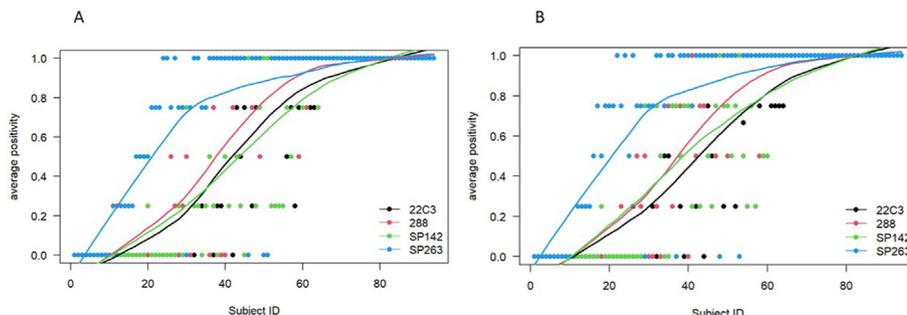
Fig. 2. PD-L1 positivity according to IC-score (A) and CPS (B) for each case and assay averaged over all readers.

Table 2
ICC values for agreement between pair-wise assay combinations across all readers.

Assay pair	ICC for PD-L1- IC-score positivity (95% CI)	ICC for PD-L1-CPS-positivity (95% CI)	ICC for PD-L1-TPS-positivity (95% CI)
28-8 vs. 22C3	0.618 (0.556—0.672)	0.842 (0.811—0.868)	0.878 (0.857—0.896)
28-8 vs. SP142	0.610 (0.551—0.662)	0.559 (0.495—0.616)	0.324 (0.241—0.401)
28-8 vs. SP263	0.357 (0.250—0.450)	0.520 (0.421—0.601)	0.557 (0.496—0.613)
22C3 vs. SP142	0.613 (0.558—0.662)	0.559 (0.499—0.613)	0.363 (0.285—0.436)
22C3 vs. SP263	0.314 (0.180—0.427)	0.476 (0.331—0.584)	0.586 (0.518—0.645)
SP142 vs. SP263	0.269 (0.154—0.370)	0.271 (0.160—0.369)	0.131 (0.051—0.209)

Table 3
Kappa statistics for comparison of pair-wise assay combinations across all readers.

Assay pair	IC ≥ 1% positivity (95% CI)	CPS ≥1 positivity (95% CI)	CPS ≥10 positivity (95% CI)	TPS ≥1% positivity (95% CI)
28-8 vs. 22C3	0.705 (0.604—0.805)	0.693 (0.593—0.794)	0.681 (0.582—0.78)	0.736 (0.635—0.837)
28-8 vs. SP142	0.605 (0.505—0.705)	0.665 (0.565—0.766)	0.486 (0.388—0.585)	0.386 (0.305—0.467)
28-8 vs. SP263	0.488 (0.393—0.584)	0.473 (0.378—0.568)	0.5 (0.403—0.597)	0.549 (0.452—0.646)
22C3 vs. SP142	0.645 (0.543—0.746)	0.669 (0.568—0.77)	0.577 (0.476—0.678)	0.43 (0.343—0.517)
22C3 vs. SP263	0.423 (0.332—0.514)	0.416 (0.326—0.506)	0.44 (0.349—0.53)	0.511 (0.418—0.605)
SP142 vs. SP263	0.39 (0.3—0.481)	0.442 (0.35—0.534)	0.411 (0.323—0.5)	0.201 (0.135—0.268

(good to excellent inter-reader agreement) and highest for 22C3. For TPS, the ICCs were 0.665—0.830 (good to excellent inter-reader agreement) and highest for 22C3. ICC values are given in Table 4.

We further performed kappa statistics to test the inter-reader reliability at specific cut-offs (Table 5). At the IC-score ≥1% cut-off, kappa values were 0.728—0.777 corresponding to a good to excellent agreement across readers for each assay. At the TPS ≥1% cut-off, the agreement across readers for each assay was fair to good (kappa 0.530—0.722). At the CPS ≥1 and ≥ 10 cut-off, kappa values were 0.680—0.735 and 0.52—0.671, respectively, again pointing on a good agreement across readers for each assay.

3.4. Inter-score differences

Finally, we analyzed the agreement between IC-score and CPS in the classification of positive and negative cases. When comparing IC-score 1% with CPS 1, kappa values were ≥0.70 for all assays indicating a good to excellent agreement. The highest kappa values (≥0.83) were identified for the SP263 assay. In contrast, kappa values dropped substantially to 0.17—0.54 when assessing the agreement between IC-score ≥1% and CPS ≥10 cut-off. The percentage overlap of positive and negative cases between IC-score ≥1% and CPS ≥1 and ≥ 10 cut-off, respectively for each assay averaged over all readers is illustrated in Fig. 3A and B. The comparison shows a concordance above 90% for PD-L1-IC-positivity ≥1% and CPS ≥1 for each assay. In contrast, comparing PD-L1-IC-positivity ≥1% and CPS ≥10 cut-off for each assay, the concordance level is less than 90%.

4. Discussion

Our comparison study of four PD-L1 assays in primary TNBC revealed a comparable PD-L1 expression on IC with SP142, 22C3, and 28—8, but higher values with SP263 as observed in our previous study [10]. For SP142 IC-score 1%, we found a mean case positivity rate of 53% that is slightly higher as compared to 46% of the biomarker-evaluable population from the pivotal IMpassion130 trial [16] but almost in line with 58% of another study [17]. Whereas substantially higher case positivity rates of 75% for SP263 are completely consistent with post-hoc biomarker evaluation data from this trial [16] and another study [17]. However, a considerably higher prevalence of PD-L1 (≥75%) with 22C3 was reported in clinical trials and exploratory biomarker analyses [5,7,16]. Nevertheless, comparable expression levels (around 60%) to our observation with 22C3 and 28—8 at CPS 1 were reported in another TNBC study [18]. Differences in PD-L1 prevalence may be explained by type of biomaterial evaluated. In our study, whole tissue sections of primary tumor resection specimens of untreated patients were used to minimize sample bias. Clinical trials as mentioned above use a mix of biopsies and resection specimens. Thus, in a certain number of cases less tumor tissue is available which might result in differing positivity rates. In this situation, we speculate that pathologists may tend to score a sample as positive, just not to miss a positive case. Conversely, a higher tumor amount offers more confidence in categorization of the given cut-offs. Furthermore, in clinical trials, tissue samples not only from primaries but also from metastatic sites were included. As reported, heterogeneity in PD-L1 expression between primary TNBC and metastatic lesions is evident [19].

In our analysis the PD-L1 positivity rate at IC-score 1% was comparable with CPS 1. In contrast to the IC-score, the CPS includes both positive IC and TC, we therefore separately analyzed TC (reported as TPS) to get an idea about the impact of TC positivity on CPS in TNBC. According to the staining pattern that is known for the different assays, we herein reproduced lower TPS values with SP142 as compared to the other assays [1,18,20]. For SP142, we found that the TPS did not strongly influence CPS at low cut-off because concomitant immune cell positivity already covered most

Table 4
ICC values for inter-reader agreement for each assay based on four readers.

Assay	ICC for PD-L1- IC-score positivity (95% CI)	ICC for PD-L1-CPS-positivity (95% CI)	ICC for PD-L1-TPS-positivity (95% CI)
SP142	0.793 (0.739—0.839)	0.720 (0.650—0.782)	0.764 (0.709—0.814)
SP263	0.489 (0.384—0.588)	0.653 (0.519—0.750)	0.665 (0.578—0.740)
22C3	0.656 (0.570—0.732)	0.794 (0.737—0.842)	0.830 (0.787—0.868)
28—8	0.537 (0.434—0.630)	0.736 (0.662—0.797)	0.730 (0.668—0.786)

Table 5
Inter-reader reliability (Kappa statistics).

Assay	IC ≥ 1% positivity (95% CI)	CPS ≥1 positivity (95% CI)	CPS ≥10 positivity (95% CI)	TPS ≥1% positivity (95% CI)
SP142	0.728 (0.645–0.811)	0.68 (0.597–0.762)	0.671 (0.589–0.754)	0.53 (0.448–0.613)
SP263	0.736 (0.653–0.819)	0.735 (0.652–0.818)	0.624 (0.541–0.707)	0.614 (0.531–0.697)
22C3	0.736 (0.653–0.819)	0.696 (0.613–0.779)	0.642 (0.559–0.725)	0.722 (0.639–0.805)
28–8	0.777 (0.694–0.86)	0.734 (0.651–0.817)	0.52 (0.437–0.603)	0.654 (0.571–0.737)
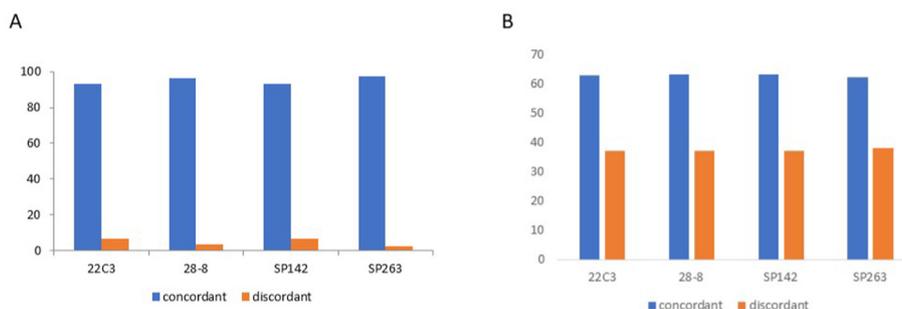
Fig. 3. Concordance between IC-score ≥ 1% and CPS ≥1 (A) and between IC-score ≥ 1% and CPS ≥10 (B) across all readers for each assay.

of the SP142 TPS negative cases. Similarly, the higher TPS with 28–8 did not generate more CPS positive cases when compared to 22C3 because these cases were already covered by positive immune cell scoring.

In our study, we observed comparable positivity rates and a good agreement between 22C3 and 28–8, a slightly less agreement for SP142 and a considerably less agreement for SP263 in CPS evaluation. Analytical performance of these four assays was previously assessed in TNBC by a single pathologist study [18]. In this abstract, a low concordance between SP263 and SP142 but a high concordance between SP263 with the other assays at CPS 1 and IC 1% was shown. Moreover, a lower prevalence of positive cases (around 30%) with SP142 was noted, resulting in almost 20% fewer PD-L1 positive TNBC with SP142 as compared to 22C3. Likewise, in a sub-study of IMpassion130, 22C3 detected almost 30% more PD-L1 positive TNBC as compared to SP142 when the IC 1% cut-off was applied [16]. In a tissue microarray study with TNBC, an even lower IC positivity with SP142 (19.3%) as compared to 28–8 (36.7%) was observed resulting in a discordance rate of 24.8% [20].

Taken all studies together, there is undoubtedly a variation between PD-L1 IHC assays that might depend on protocols, detection kits and platforms used as discussed earlier [21]. In addition, primary PD-L1 antibodies show certain differences between TC and IC staining patterns, e. g. SP142 was developed and optimized for IC detection while SP263 detects more TC [13,22]. In a linear epitope mapping experiment, different binding characteristics for PD-L1 antibodies were found that could lead to staining discrepancies [23]. As outlined above, material used (TMA versus whole tissue sections or biopsies versus surgical resection specimens) has undoubtedly a certain impact on assay concordance, especially since the staining of consecutive sections from small amounts of tissue (like in the biopsy or TMA situation) may overestimate discrepancies, because variations in positivity rates based on tumor heterogeneity might easily be misinterpreted as a difference in assay performance. Furthermore, qualified pathologists for the different scoring algorithms and staining properties of different antibodies are key for a correct PD-L1 assessment. All readers in our study

were trained for CPS and IC evaluation in training programs offered by Roche as well as Merck, Sharp & Dome. In addition, all have ample experience with comparability studies for PD-L1 assays in a variety of tumor entities including the same assays [10,24–26]. Finally, in an assay comparability study, it is crucial that all cases are evaluated by the same reader to avoid interobserver bias, especially if the readers are not perfectly trained. In our study, we also assessed inter-reader concordance and demonstrate an overall good reproducibility for all scorings and each assay among the pathologists that is in line with previous PD-L1 comparability studies [27,28]. Collectively, the variation of PD-L1 prevalence between different studies needs further evaluation to better understand the inter-assay differences. To overcome the challenge of predictive PD-L1 testing, recommendations that guide this complex issue are very useful [29].

To date, PD-L1 is the most widely used biomarker for patient selection in ICI treatment. However, not all patients with positive PD-L1 status respond to these therapies. Attempts to define other biomarkers (e. g. TILs, CD8 cells, tumor mutational burden, BRCA) that provide additional value beyond PD-L1 in patient selection in TNBC failed so far [1], yet data from other tumor entities and pan-cancer studies implicated that additional molecular markers such as microsatellite instability (MSI) and high tumor mutational burden (TMB) can be successfully applied [30,31].

Taken together, we demonstrate comparable positivity rates for 22C3, 28–8, and SP142 with IC 1%, accompanied by an overall good assay concordance and without an antibody dependent specific bias in one direction. Our data may suggest that 22C3, 28–8, and SP142 can be used for IC evaluation in routine diagnostics like in other entities such as urothelial cancer [24]. Further confirmation by other studies is needed. 22C3 and 28–8 can likely be used interchangeable for IC 1% and CPS 1, but less with CPS 10. SP142 showed still good concordance for IC 1% and CPS 1 but less overlap with CPS 10, this in sum with the confirmed lower positivity of SP142 in tumor cells, might render SP142 a less optimal antibody for CPS assessment in TNBC, specifically at higher cut-offs. SP263 selects more positive cases with both IC and CPS and therefore in

our view should not be used for PD-L1 scoring in TNBC. Finally, interchangeability of PD-L1 assays is not only based on inter-assay concordance in terms of prevalence and positivity but also on the capacity of each assay to predict response to a given ICI [16].

Author contributions

WW, DO, AN were involved in the conception of the study. WW, AN, DO, WR contributed to the study design. The first draft of the manuscript was written by AN and WW. All authors were involved in material preparation, data collection and analysis, and in reviewing and editing the manuscript content, as well as approving the final manuscript for submission.

Source of funding

Declaration of competing interest

AN has attended Advisory Boards. DCW has received funding from Roche Pharma AG, BMS, and MSD. KSc has attended Advisory Boards and served as a speaker for Roche, MSD, and BMS. KS has received funding from Roche Pharma AG. MK has received remuneration from Springer Press, Biermann Press, Celgene, AstraZeneca, Myriad Genetics and Teva, received consultancy or advisory fees from Myriad Genetics, KVB, DKMS LIFE, BLÄK and TEVA, holds stock in Therawis Diagnostics GmbH and AIM GmbH and received funding from Sphingotec, Deutsche Krebshilfe, DFG, BMBF, the Senator Roesner Foundation and the Dr Pommer-Jung Foundation. DO is an employee of MSD Sharp & Dohme GmbH. WR has received funding from Roche Pharma AG, has attended Advisory Boards and served as a speaker for Roche, MSD, Novartis. WW has attended Advisory Boards, served as speaker for Roche, MSD, BMS, AstraZeneca, Pfizer, Merck, Lilly, Boehringer, Novartis, Takeda, Bayer, Amgen, Astellas, Illumina, Siemens, Agilent and Molecular Health and receives research funding from Roche, MSD, BMS and AstraZeneca. The remaining authors (SF, SK, AH) have no conflict of interest.

Acknowledgements

References

[1] Emens LA. Predictive biomarkers: progress on the Road to personalized cancer immunotherapy. J Natl Cancer Inst; 2021.

[2] Nagarajan D, McArdle SEB. Immune landscape of breast cancers. Biomedicines 2018;6(1).

[3] Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. N Engl J Med 2018;379(22):2108—21.

[4] Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 2020;21(1):44—59.

[5] Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. Lancet 2020;396(10265):1817—28.

[6] Mittendorf EA, Zhang H, Barrios CH, Saji S, Jung KH, Hegg R, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. Lancet 2020;396(10257):1090—100.

[7] Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. N Engl J Med 2020;382(9):810—21.

[8] Cardoso F, Paluch-Shimon S, Senkus E, Curigliano G, Aapro MS, André F, et al. 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). Ann Oncol 2020;31(12):1623—49.

[9] Committee AB. Diagnosis and treatment of patients with primary and metastatic breast cancer. 2021. Recommendations, www.ago-online.de.

[10] Noske A, Ammann JU, Wagner DC, Denkert C, Lebeau A, Sinn P, et al. A multicentre analytical comparison study of inter-reader and inter-assay agreement of four programmed death-ligand 1 immunohistochemistry assays for scoring in triple-negative breast cancer. Histopathology 2021;78(4):567—77.

[11] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline focused update. J Clin Oncol 2018;36(20):2105—22.

[12] Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol 2020;38(12):1346—66.

[13] Vennapusa B, Baker B, Kowanetz M, Boone J, Menzl I, Bruey JM, et al. Development of a PD-L1 complementary diagnostic immunohistochemistry assay (SP142) for atezolizumab. Appl Immunohistochem Mol Morphol 2019;27(2):92—100.

[14] Kulangara K, Zhang N, Corigliano E, Guerrero L, Waldroup S, Jaiswal D, et al. Clinical utility of the combined positive score for programmed death ligand 1 expression and the approval of pembrolizumab for treatment of gastric cancer. Arch Pathol Lab Med 2019;143(3):330—7.

[15] Cicchetti DV. Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research. J Clin Exp Neuropsychol 1994;16(1):155—61.

[16] Rugo HS, Loi S, Adams S, Schmid P, Schneeweiss A, Barrios CH, et al. PD-L1 immunohistochemistry assay comparison in atezolizumab plus nab-paclitaxel-treated advanced triple-negative breast cancer. J Natl Cancer Inst; 2021.

[17] Reisenbichler ES, Han G, Bellizzi A, Bossuyt V, Brock J, Cole K, et al. Prospective multi-institutional evaluation of pathologist assessment of PD-L1 assays for patient selection in triple negative breast cancer. Mod Pathol 2020;33(9):1746—52.

[18] Scott M, Scorer P, Barker C, Al-Masri H. Comparison of patient populations identified by different PD-L1 assays in triple-negative breast cancer (TNBC). Ann Oncol 2019;(suppl_3):30. https://doi.org/10.1093/annonc/mdz095. iii1-iii26.

[19] Rozenblit M, Huang R, Danziger N, Hegde P, Alexander B, Ramkissoon S, et al. Comparison of PD-L1 protein expression between primary tumors and metastatic lesions in triple negative breast cancers. J Immunother Cancer 2020;8(2).

[20] Sun WY, Lee YK, Koo JS. Expression of PD-L1 in triple-negative breast cancer based on different immunohistochemical antibodies. J Transl Med 2016;14(1):173.

[21] Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. JAMA Oncol 2017;3(8):1051—8.

[22] Udall M, Rizzo M, Kenny J, Doherty J, Dahm S, Robbins P, et al. PD-L1 diagnostic tests: a systematic literature review of scoring algorithms and test-validation metrics. Diagn Pathol 2018;13(1):12.

[23] Schats K, Vre EAV, Schrijvers D, Meester ID, Kockx M. Epitope mapping of PD-L1 primary antibodies (28-8, SP142, SP263, E1L3N). J Clin Oncol 2017;35(15_suppl):3028.

[24] Schwamborn K, Ammann JU, Knüchel R, Hartmann A, Baretton G, Lasitschka F, et al. Multicentric analytical comparability study of programmed death-ligand 1 expression on tumor-infiltrating immune cells and tumor cells in urothelial bladder cancer using four clinically developed immunohistochemistry assays. Virchows Arch 2019;475(5):599—608.

[25] Sommer U, Eckstein M, Ammann J, Braunschweig T, Macher-Göppinger S, Schwamborn K, et al. Multicentric analytical and inter-observer comparability of four clinically developed programmed death-ligand 1 immunohistochemistry assays in advanced clear-cell renal cell carcinoma. Clin Genitourin Cancer

2020;18(5):e629—42.

[26] Scheel AH, Baenfer G, Baretton G, Dietel M, Diezko R, Henkel T, et al. Inter-laboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. Histopathology 2018;72(3):449—59.

[27] Downes MR, Slodkowska E, Katabi N, Jungbluth AA, Xu B. Inter- and intra-observer agreement of programmed death ligand 1 scoring in head and neck squamous cell carcinoma, urothelial carcinoma and breast carcinoma. Histo-pathology 2020;76(2):191—200.

[28] Wang C, Hahn E, Slodkowska E, Eskander A, Enepekides D, Higgins K, 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—9.

[29] Cheung CC, Barnes P, Bigras G, Boerner S, Butany J, Calabrese F, et al. Fit-for-purpose PD-L1 biomarker testing for patient selection in immuno-oncology: guidelines for clinical laboratories from the Canadian association of pathologists-association canadienne des pathologistes (CAP-ACP). Appl Immunohistochem Mol Morphol 2019;27(10):699—714.

[30] Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord JP, et al. Efficacy of pembrolizumab in patients with noncolorectal high micro-satellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. J Clin Oncol 2020;38(1):1—10.

[31] Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet 2019;51(2):202—6.