




# 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

Aurelia Noske,<sup>1</sup>  Johannes U Ammann,<sup>2,\*</sup>  Daniel-Christoph Wagner,<sup>3</sup> Carsten Denkert,<sup>4</sup> Annette Lebeau,<sup>5,6</sup> Peter Sinn,<sup>7</sup>  Hans-Heinrich Kreipe,<sup>8</sup> Ulrich Sommer,<sup>9</sup> Gustavo Baretton,<sup>9</sup> Katja Steiger,<sup>1</sup> Marion Kiechle,<sup>10</sup>  Stefanie Hieke-Schulz,<sup>2</sup> Mike Flores,<sup>11</sup> Wilfried Roth<sup>3</sup> & Wilko Weichert<sup>1</sup>

<sup>1</sup>Technical University of Munich, Institute of Pathology, Munich, <sup>2</sup>Roche Pharma AG, Grenzach-Wyhlen, <sup>3</sup>Institute of Pathology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, <sup>4</sup>Institute of Pathology, Philipps-University Marburg and University Hospital of Giessen and Marburg, Marburg, <sup>5</sup>Private Group Practice for Pathology Lübeck, Lübeck, <sup>6</sup>Department of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, <sup>7</sup>Division of Gynecopathology, University Institute of Pathology, Heidelberg University Hospital, Heidelberg, <sup>8</sup>Institute of Pathology, Hannover Medical School, Hannover, <sup>9</sup>Institute of Pathology, University Hospital Carl Gustav Carus Dresden, Technical University of Dresden, Dresden, <sup>10</sup>Department of Gynecology and Obstetrics, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany, and <sup>11</sup>Ventana Medical Systems, Inc., Tucson, AZ, USA

Date of submission 24 March 2020

Accepted for publication 9 September 2020

Published online Article Accepted 16 September 2020

Noske A, Ammann J U, Wagner D-C, Denkert C, Lebeau A, Sinn P, Kreipe H-H, Sommer U, Baretton G, Steiger K, Kiechle M, Hieke-Schulz S, Flores M, Roth W & Weichert W

(2021) *Histopathology* 78, 567–577. <https://doi.org/10.1111/his.14254>

## 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

**Aims:** Studies in various cancer types have demonstrated discordance between results from different programmed death-ligand 1 (PD-L1) assays. Here, we compare the reproducibility and analytical concordance of four clinically developed assays for assessing PD-L1-positivity in tumour-infiltrating immune cells in the tumour area (PD-L1-IC-positivity) in triple-negative breast cancer (TNBC).

**Methods and results:** Primary TNBC resection specimens ( $n = 30$ ) were selected based on their PD-L1-IC-positivity per VENTANA SP142 (<1%: 15 cases; 1–5%: seven cases; >5%: eight cases). Serial histological sections were stained for PD-L1 using VENTANA

SP142, VENTANA SP263, DAKO 22C3 and DAKO 28-8. PD-L1-IC-positivity and tumour cell expression ( $\geq 1$  versus <1%) were scored by trained readers from seven sites using online virtual microscopy. The adjusted mean of PD-L1-IC-positivity for SP263 (7.8%) was significantly higher than those for the other three assays (3.7–4.9%). Differences in adjusted means were statistically significant between SP263 and the other three assays ( $P < 0.0001$ ) but not between the three remaining assays when excluding SP263 ( $P = 0.0961$ – $0.6522$ ). Intra-class correlation coefficients revealed moderate-to-strong inter-reader agreement for each assay (0.460–0.805) and poor-

Address for correspondence: A Noske, Technical University of Munich, School of Medicine, Institute of Pathology, Trogerstrasse 18, 81675 Munich, Germany. e-mail: aurelia.noske@tum.de

F. Hoffmann-La Roche Ltd, Basel, Switzerland

\*Present address: F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland.

A Noske and J U Ammann contributed equally to this work.

© 2020 The Authors. *Histopathology* published by John Wiley & Sons Ltd

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

to-strong inter-assay agreement for each reader (0.298–0.678) on PD-L1-IC-positivity.

**Conclusions:** In this first multicentre study of different PD-L1 assays in TNBC, we show that PD-L1-IC-positivity for SP142, 22C3 and 28-8 was

**Keywords:** immunohistochemistry, inter-assay agreement, inter-reader agreement, programmed death-ligand 1, triple-negative breast cancer

## Introduction

Triple-negative breast cancer (TNBC) lacks oestrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER2) overexpression/amplification, is characterised by fast disease progression and has limited systemic therapy regimens and lower survival rates versus other BCs.<sup>1–6</sup>

For patients with unresectable locally advanced/metastatic TNBC (mTNBC) who have programmed death-ligand 1 (PD-L1)-stained tumour-infiltrating immune cells (IC) covering  $\geq 1\%$  of the tumour area, atezolizumab plus nab-paclitaxel was approved as a first-line option by the Food and Drug Administration (FDA) and European Medicines Agency (EMA),<sup>7,8</sup> and is recommended by National Comprehensive Cancer Network and Arbeitsgemeinschaft Gynäkologische Onkologie guidelines.<sup>3,6</sup> Compared with the FDA label, the EMA label does not restrict the PD-L1 assay that can be used for PD-L1 IC status assessment.

Atezolizumab first-line monotherapy in mTNBC showed a more pronounced response in patients with higher IC levels and PD-L1-IC-positivity.<sup>9</sup> Adding chemotherapy to atezolizumab may have a synergistic effect, enhancing anti-tumour activity.<sup>10</sup>

IMpassion130 demonstrated a significant 2.5-month progression-free survival (PFS) benefit (hazard ratio [HR] = 0.62; 95% confidence interval [CI] = 0.49–0.78;  $P < 0.001$ ) with atezolizumab plus nab-paclitaxel versus placebo plus nab-paclitaxel in patients with PD-L1-IC-positive TNBC.<sup>11</sup> A clinically meaningful overall survival (OS) improvement of 7.0 months (HR = 0.71; 95% CI = 0.54–0.93) was also demonstrated.<sup>12</sup> PFS/OS improvements in the intention-to-treat population were smaller, and driven by the PD-L1-IC-positive population.<sup>11–13</sup> These results, and the FDA and EMA approval, underline the importance of patient selection based on testing PD-L1-IC status.

Immune checkpoint proteins, such as those of the PD-L1/programmed cell death-1 (PD-1) pathway, block development of active anti-tumour immune responses. PD-L1 is expressed on tumour cells (TC) and IC, among others.<sup>10,14</sup> Several PD-L1

reproducible and analytically concordant, indicating that these three assays may be analytically interchangeable. The relevance of the higher PD-L1-IC-positivity for SP263 should be further investigated.

immunohistochemistry (IHC) assays are under development/have been approved as companion/complementary diagnostics to anti-PD-L1/PD-1 therapies. Four have been used extensively in clinical trials; VENTANA SP142 (Roche Diagnostics, Mannheim, Germany) with atezolizumab; VENTANA SP263 (Roche Diagnostics) with durvalumab; DAKO 22C3 (Agilent Technologies, Waldbronn, Germany) with pembrolizumab; and DAKO 28-8 (Agilent Technologies) with nivolumab (Table S1).

Analytical discordance between different assays has been reported in 25–50% of samples from patients with non-small-cell lung cancer; PD-L1 TC staining being reduced with SP142 versus SP263, 22C3 and 28-8.<sup>15–18</sup> SP142, SP263, 22C3 and 28-8 give similar PD-L1-IC-positivities, with moderate-to-strong inter-assay and inter-reader agreement in urothelial and clear cell renal cell carcinoma.<sup>19–24</sup> Published data on analytical concordance of PD-L1-IC assays in TNBC are limited. We aimed to compare reproducibility and analytical concordance of four PD-L1 IHC assays used in registrational trials with trained readers focusing on PD-L1-IC scoring in a real-world setting. We report results from this first multicentre PD-L1-IC comparability study, based on staining of whole slides, in TNBC.

## Materials and methods

### STUDY DESIGN

This was a multicentre biomarker study, designed to assess analytical and inter-reader comparability of PD-L1-IC-positivity across four PD-L1 IHC assays in TNBC. One hundred and seven archival, formalin-fixed, paraffin-embedded, primary TNBC resection specimens from the Technical University of Munich (TUM) were randomly selected and screened for PD-L1-IC-positivity using SP142, as per the manufacturer's protocol. Samples were processed between 2003 and 2014, and selected according to hormone receptor- and HER2-negativity. Specimens needed enough tissue per block to allow production of  $\geq 10$  serial slides. From these, 30 cases were selected to give

a range of PD-L1-IC-positivity, which included a representative number of cases in all scoring categories used in pivotal atezolizumab studies.<sup>13,25</sup> The final selection included 15 cases with PD-L1-IC-positivity of <1% of tumour area, seven cases with 1–5% and eight cases with >5%. 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). Data S1 provides the IHC details. Prior to assessment, all readers with PD-L1 testing expertise across cancer types/assays were trained on PD-L1-IC staining with SP142,<sup>26</sup> using 75 cases across a dynamic range of PD-L1-positivity scored according to the SP142 algorithm on PathoTrainer (Pathomation Inc., Antwerp, Belgium), and had to pass a 40-case proficiency examination (minimum score = 85%). Readers were not trained on PD-L1-TC scoring. Trained readers at seven sites were given access to sets of slides stained with the PD-L1 assays and their corresponding haematoxylin and eosin (H&E)/pan-cytokeratin stains, which were randomised and blinded for patient and assay information on an online digital platform by excluding the on-slide positive controls from scanning (following visual inspection of control for PD-L1-positivity) and giving each slide a dedicated ID (1–120). PD-L1-stained slides were scored for PD-L1-IC-positivity (percentage of tumour area covered by stained IC) and PD-L1-TC expression (percentage of stained TC in the tumour area;  $\geq 1\%$  versus <1%) by all readers. SP142-stained slides were also scored for PD-L1-IC-positivity by an additional expert reader (Reader 8) from Ventana Medical Systems, Inc. PD-L1-IC-positivity was defined as staining in granulocytes, lymphocytes, macrophages and dendritic cells of any intensity within the tumour area. PD-L1-TC-positivity was defined as membranous PD-L1 staining of any intensity in  $\geq 1\%$  of TC.<sup>27</sup> Reading times were measured using web browser activity tracking. Presence of tumour-infiltrating lymphocytes (TILs) was evaluated by A.N., as published.<sup>28</sup> 22C3-stained slides were additionally read for PD-L1 combined positivity score (CPS) by U.S., who was trained across cancer types by the manufacturer (Data S2).

#### OBJECTIVES

The primary objective was to compare PD-L1-IC-positivity across assays, adjusted for reader effects. The secondary objective was to assess inter-reader agreement on PD-L1-IC-positivity for each assay, inter-assay agreement on PD-L1-IC-positivity for each reader, percentage of inter-reader agreement for assay pairs at the  $\geq 1\%$  cut-off and differences in reading times for

scoring PD-L1-IC-positivity between assays for each reader and between readers for each assay. Additional objectives were exploration of inter-assay differences in PD-L1-TC-positivity, investigation of the distribution of TILs and exploration of the association between PD-L1-IC-positivity (per SP142) and the CPS (per 22C3). Data S3 shows the statistical methods.

Ethical approval was not required for this study. The experiments complied with the current laws of the country in which they were performed. Informed consent for tissue analysis was obtained from all individuals whose TNBC tissue specimens were included in this study.

## Results

#### PD-L1-IC-POSITIVITY (SCREENING AND STUDY COHORTS)

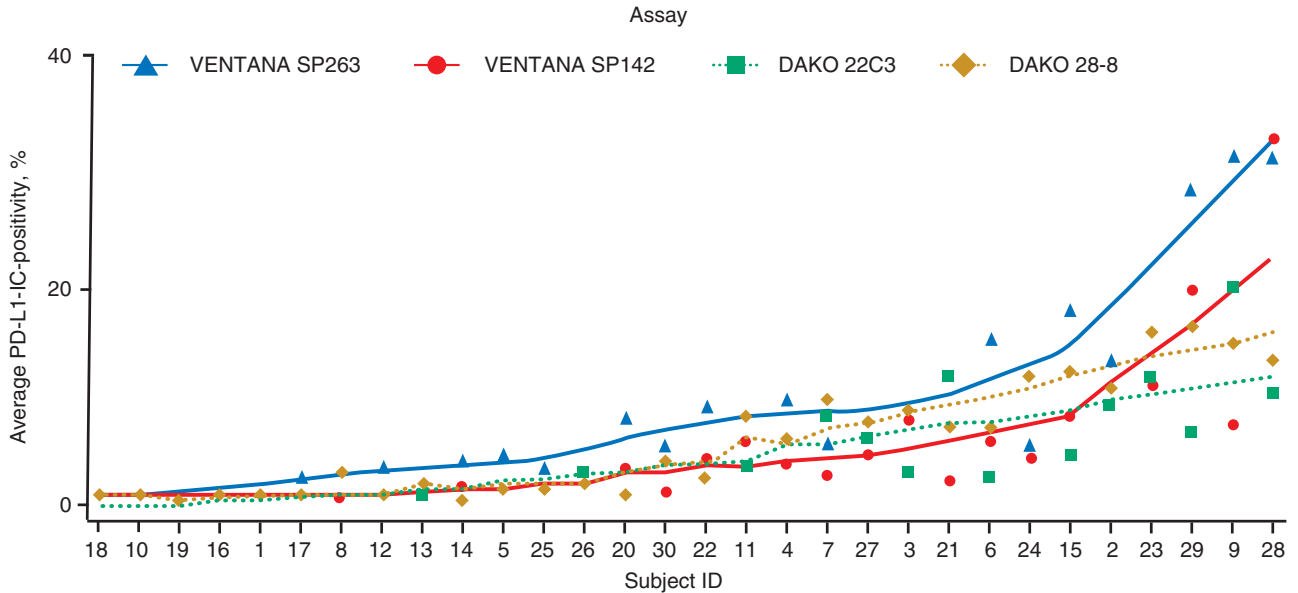
A total of 104 of 107 samples in the screening cohort were evaluable. Distributions of PD-L1-IC-positivity per SP142 in the screening and study cohorts ( $n = 30$ ) were similar (Table S2). Table S3 shows clinicopathological characteristics of the study cohort.

#### BETWEEN-ASSAY PD-L1-IC-POSITIVITY COMPARISON (PRIMARY OBJECTIVE)

Adjusted mean PD-L1-IC-positivity was similar for SP142, 22C3 and 28-8 (range = 3.7–4.9%), but was significantly higher for SP263 (7.8%) (Figure 1; Table 1). Differences between adjusted means of PD-L1-IC-positivity for SP142, 22C3 and 28-8 were  $-1.2$  to  $-0.6\%$  ( $P = 0.0961$ – $0.6522$ ) (Figure 2). Differences in adjusted means between SP263 and the other assays were 3.0–4.2% ( $P < 0.0001$ ) (Figure 2). *Post-hoc* analyses in the screening cohort confirmed a higher proportion of patients with PD-L1-IC-positivity  $\geq 1\%$  with SP263 versus SP142 (Table S2). Additional *post-hoc* analyses in the study cohort comparing staining patterns in samples with PD-L1-TC-positivity <1% and  $\geq 1\%$  confirmed the observed higher PD-L1-IC-positivity with SP263 in both subgroups.

#### INTER-READER AND PER-READER INTER-ASSAY PD-L1-IC-POSITIVITY AGREEMENT

Intra-class correlation coefficients (ICCs) for each assay were 0.460–0.805 (moderate-to-strong inter-reader agreement) and the highest, with SP142 (Table 2). When adjusted mean PD-L1-IC-positivity scores for readers were compared with those of the expert, scores for six of seven (2.60–5.17%) readers were not



**Figure 1.** Percentage of PD-L1-IC-positivity using each assay (averaged over the seven readers). IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1.

**Table 1.** Adjusted mean percentages of PD-L1-IC-positivity across all samples using each assay, adjusted for reader and sample effects

Assay	Adjusted mean PD-L1-IC-positivity, % (95% CI)
VENTANA SP263	7.8 (7.1–8.6)
VENTANA SP142	4.3 (3.5–5.0)
DAKO 22C3	3.7 (2.9–4.4)
DAKO 28-8	4.9 (4.1–5.6)

CI, Confidence interval; IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1.

statistically significantly different (4.00%). For the remaining reader, adjusted mean PD-L1-IC-positivity was 6.38%. Figure S1 shows differences in adjusted means of PD-L1-IC-positivity for each reader compared with the expert. ICCs for each reader were 0.298–0.678 (poor-to-strong inter-assay agreement for each) (Table 2).

AGREEMENT ON PD-L1-IC-POSITIVITY STATUS AT THE ≥1% CUT-OFF

Figure 3 shows heat-maps<sup>29</sup> for PD-L1-IC-positivity at the ≥1% cut-off. Concordance for each sample was similar for SP142, 22C3 and 28-8; only four cases

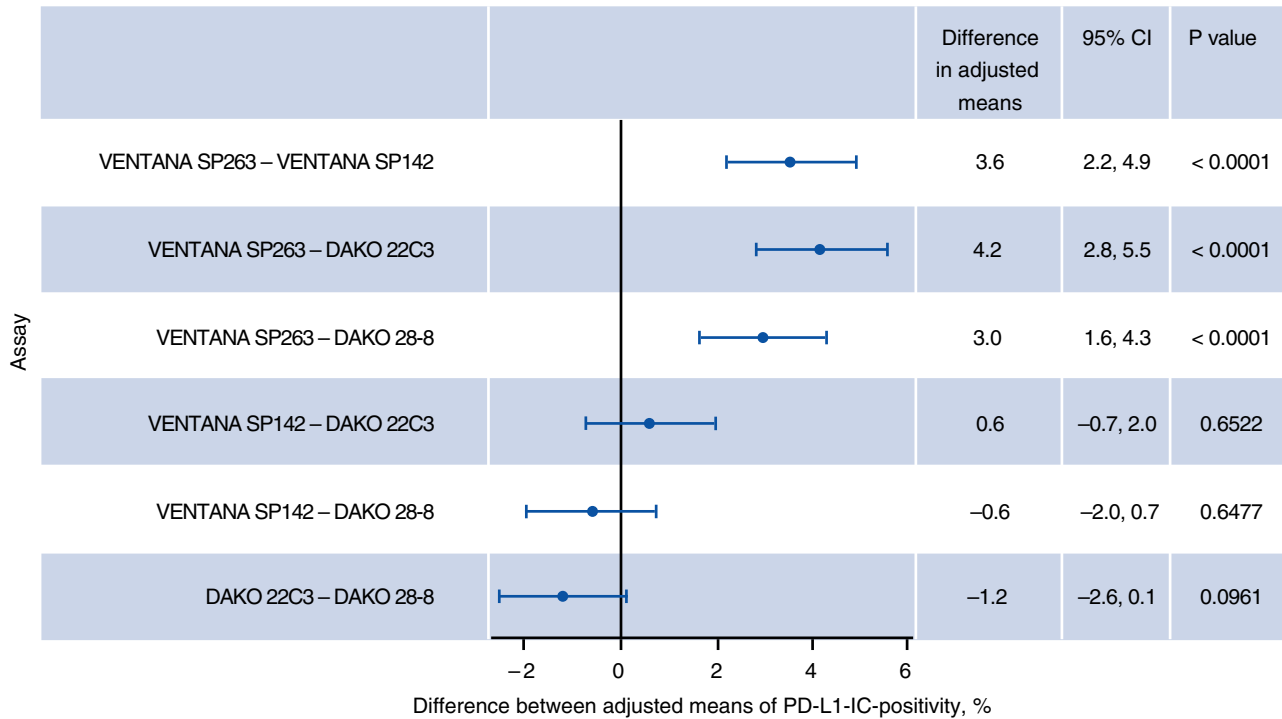
had discrepant results. With SP263, discrepant results were observed in an additional six cases (Figure 3). At the ≥1% cut-off, kappa scores were 0.589–0.789 (moderate-to-strong concordance across readers for each assay) (Table S4). Concordance across assays for each reader at the ≥1% cut-off was fair-to-strong (kappa scores 0.358–0.704; Table S5).

Similar results were observed when the average percentage agreements for PD-L1-IC-positivity at the ≥1% cut-off were assessed between all assay pair combinations (Figure 4). Disagreement was observed in 10% of cases for SP142 versus 22C3 or 28-8 (Figure 4); disagreement was observed in 23–30% for SP263 versus the other assays (Figure 4).

Figure S2 shows agreement on PD-L1-TC-positivity of ≥1% across readers for each of the samples (columns) for each of the assays (Data S4 provides additional information).

DIFFERENCES IN READING TIMES FOR PD-L1-IC-POSITIVITY

Reading times were assessed for six of seven readers. One was excluded as their time was not recorded due to using an old browser. A trend for faster SP142 PD-L1-IC reading times was observed for five of six (1, 2, 4, 6 and 7) (Figure 5). However, differences in times between assays were marginal for most readers (Figure 5).



**Figure 2.** Differences in adjusted means of PD-L1-IC-positivity for each assay; 0 indicates no difference between adjusted means of PD-L1-IC-positivity. Differences between adjusted means (including 95% CIs) were based on an ANOVA model with sample, assay and reader as considered effects. ANOVA, Analysis of variance; CI, Confidence interval; IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1.

#### DISTRIBUTION OF TILS IN SCREENING AND STUDY COHORTS

Distributions of TILs in the screening and study cohorts were similar (Table S6). Although TILs were present in almost all TNBC tissue specimens [ $n = 100$  (96.2%)] in the screening cohort and all 30 cases in the study cohort (100%), specimens with high TIL levels ( $\geq 60\%$ ) were rare [screening cohort: 1 (1%); study cohort: 0 (0%)] (Table S6). In the screening cohort, there appeared to be an association between TILs and PD-L1-IC-positivity using the  $\geq 30\%$  TIL cut-off, with an increasing proportion of samples with increasing PD-L1-IC-positivity per SP142 and SP263 for  $\geq 30\%$  TILs and vice versa for  $< 30\%$  TILs (Table S7).

#### OVERLAP BETWEEN PD-L1-IC- AND CPS-POSITIVITY

Table S8 shows overlap of samples with SP142 PD-L1-IC-positivity  $\geq 1\%$  and samples with 22C3 CPS  $\geq 1$  (assessed by one reader). Overall, 22 patients had PD-L1-IC-positivity  $\geq 1\%$  with SP142 and 16 had CPS  $\geq 1$  with 22C3. Cross-comparisons of PD-L1-IC-positivity

and CPS-positivity indicated differing PD-L1 status in 10 of 30 patients. Of these, eight were classified as PD-L1-IC-positive/CPS-negative and two as PD-L1-IC-negative/CPS-positive.

#### Discussion

In this first multicentre analysis comparing PD-L1 assays in TNBC, PD-L1-IC-positivity was similar for SP142, 22C3 and 28-8, but was significantly higher for SP263. Pre-defined, retrospective assessment of assay pairs at the PD-L1-IC  $\geq 1\%$  cut-off indicated 10% disagreement between SP142 and each DAKO assay. This was consistent among all readers, indicated by fair-to-strong concordance.

Limited partially contradictory data have been published on the staining properties of the assessed assays in TNBC. PD-L1 expression was investigated with SP263 in a large breast cancer cohort with different molecular subtypes;<sup>30</sup> 40% were PD-L1-IC-positive and 10% were PD-L1-TC-positive in the TNBC subgroup. In a study evaluating PD-L1-IC- and -TC-positivity in TNBC, discordance was observed between 28-8 and SP142, with higher expression levels with

**Table 2.** ICCs for inter-reader agreement for each assay and inter-assay agreement for each reader for PD-L1-IC-positivity

	ICC for PD-L1-IC-positivity (95% CI)
Inter-reader agreement for each assay (based on the seven readers)	
VENTANA SP263	0.616 (0.477–0.758)
VENTANA SP142	0.805 (0.710–0.887)*
DAKO 22C3	0.605 (0.474–0.755)
DAKO 28-8	0.460 (0.319–0.636)
Inter-assay agreement for each reader	
R1	0.678 (0.525–0.811)
R2	0.678 (0.526–0.811)
R3	0.641 (0.478–0.785)
R4	0.669 (0.518–0.807)
R5	0.341 (0.159–0.549)
R6	0.298 (0.119–0.510)
R7	0.477 (0.294–0.670)

CI, Confidence interval; IC, Tumour-infiltrating immune cell in the tumour area; ICC, Intra-class correlation coefficient (with reader as fixed effect); PD-L1, Programmed death-ligand 1; R, Reader.

\*When the expert reader was included, the ICC for VENTANA SP142 was 0.826 (95% CI = 0.740–0.899).

28-8 versus SP142.<sup>31</sup> Another study comparing identification of 196 patients with PD-L1-IC-positive TNBC at the  $\geq 1\%$  cut-off across the four assays by a single pathologist suggested that SP142 identified approximately 20% fewer patients than the others.<sup>32</sup> A further investigation of analytical comparability showed that 22C3 stained 29% more samples PD-L1-IC-positive versus SP142, whereas SP142 stained 2% of samples PD-L1-IC-positive, which were PD-L1-IC-negative with 22C3.<sup>33</sup> For the same sample cohort, SP263 stained 30% more samples PD-L1-IC-positive versus SP142, whereas SP142 stained 1% of samples PD-L1-IC-positive, which were PD-L1-IC-negative with SP263.<sup>34</sup> Similarly, our study identified six PD-L1-IC-positive samples with SP263, which were negative with SP142. *Post-hoc* analyses showed that in PD-L1-TC-positivity ( $< 1\%$  or  $\geq 1\%$ ) subgroups, the higher PD-L1-IC-positivity status with SP263 was confirmed, excluding the possibility that TC were mistakenly scored as IC. Furthermore, a higher percentage of PD-L1-IC-positivity  $> 1\%$  was confirmed for SP263 versus SP142 in the screening cohort, making

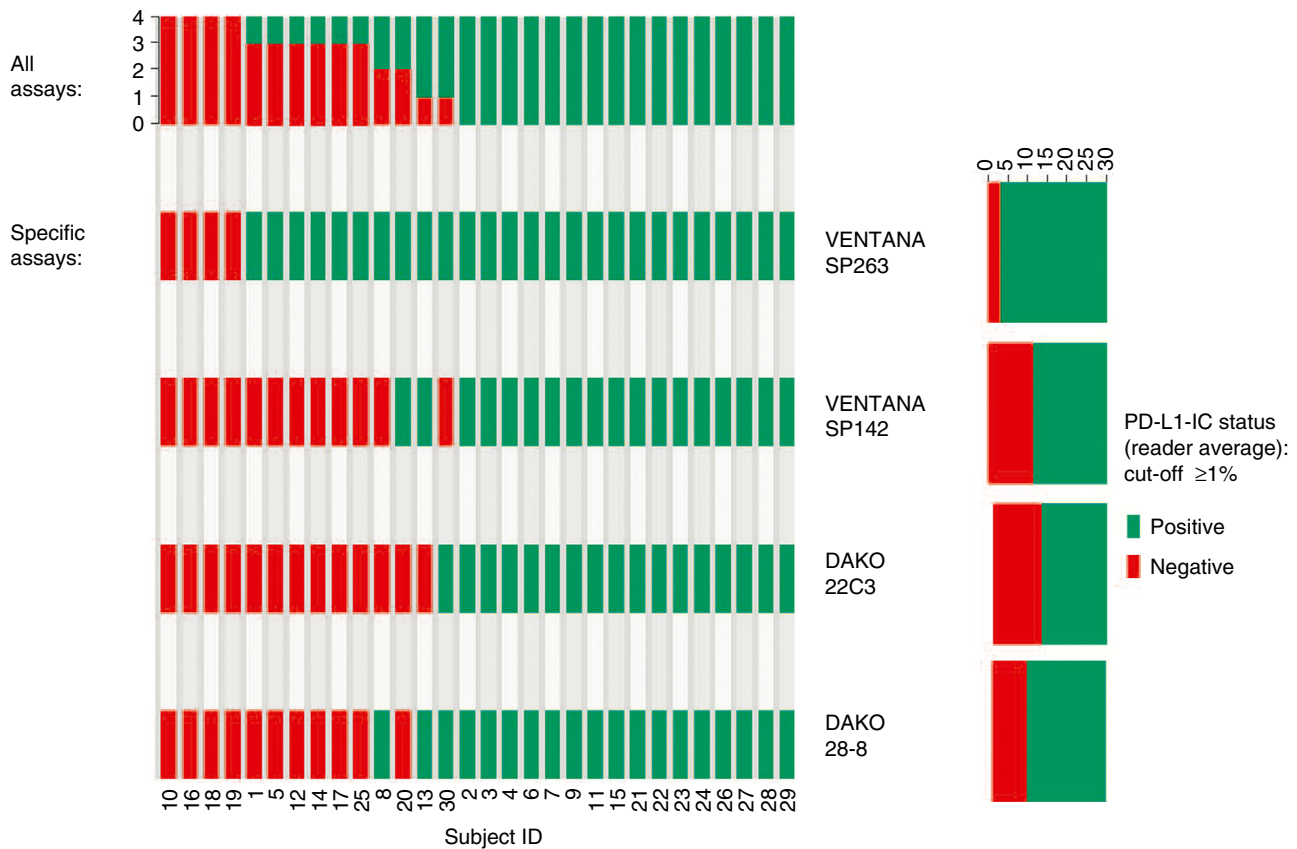
sampling bias unlikely. All evidence suggests that SP263 stains more samples PD-L1-IC-positive than SP142 in TNBC.

In our study, PD-L1-IC-positivity with 22C3 and 28-8 was more similar to that seen with SP142 than to that seen with SP263. In contrast, another study in TNBC suggested that PD-L1-IC-positivity rates with SP263 and 22C3 were similar.<sup>33,34</sup> Discrepancies between studies could be due to tumour heterogeneity or untrained readers. Several studies based on tissue microarrays have reported lower PD-L1-IC sensitivity for SP142, which may reflect sampling bias and intra-tumour heterogeneity.<sup>35,36</sup> We accounted for heterogeneity by using whole slides. To prevent difficulties in differentiating TC from IC and to outline the tumour area, both of which are key for accurately reading PD-L1-IC-positivity, every PD-L1 stain had a corresponding pan-cytokeratin stain. Another explanation for between-study discrepancies is that PD-L1-IC evaluation is more complicated than PD-L1-TC-positivity assessment. This is reflected by early lung cancer data, which demonstrated low reproducibility between readers when scoring PD-L1-IC-positivity<sup>15,17,37</sup> and lower concordance between reader scores and scores from quantitative immunofluorescence methods for PD-L1-IC-positive versus PD-L1-TC-positive.<sup>35</sup> Assay comparison studies only produce reliable results when performed by experienced, trained readers to avoid mixing inter-assay and intra-/inter-observer variability. Therefore, all readers were trained beforehand on reading PD-L1-IC-positivity and had ample experience with all staining modalities.<sup>19,24</sup> To minimise reader bias, PD-L1-IC scores were read independently by seven readers and results were adjusted for reader effects. Therefore, our results were obtained in a setting that aimed for high reproducibility and was representative of clinical practice.

Reading time analyses indicated a trend for marginally faster PD-L1-IC-positivity score reading with SP142 for five of six readers. This may be due to SP142 staining TC more weakly than the other assays. Hence, SP142 might allow for marginally quicker identification and assessment of PD-L1-IC-positivity in the tumour area. Faster reading times with SP142 may also have been influenced by the initial training for PD-L1-IC scoring with this assay.

Of particular importance for clinical practice is our comparison of PD-L1-IC- and CPS-positivities, which indicated discrepancies in classifications in one-third of patients, with more classified as PD-L1-IC-positive with SP142 than CPS-positive with 22C3. In contrast, another study showed that 22C3 identified 81% of samples as CPS-positive, while SP142 classified



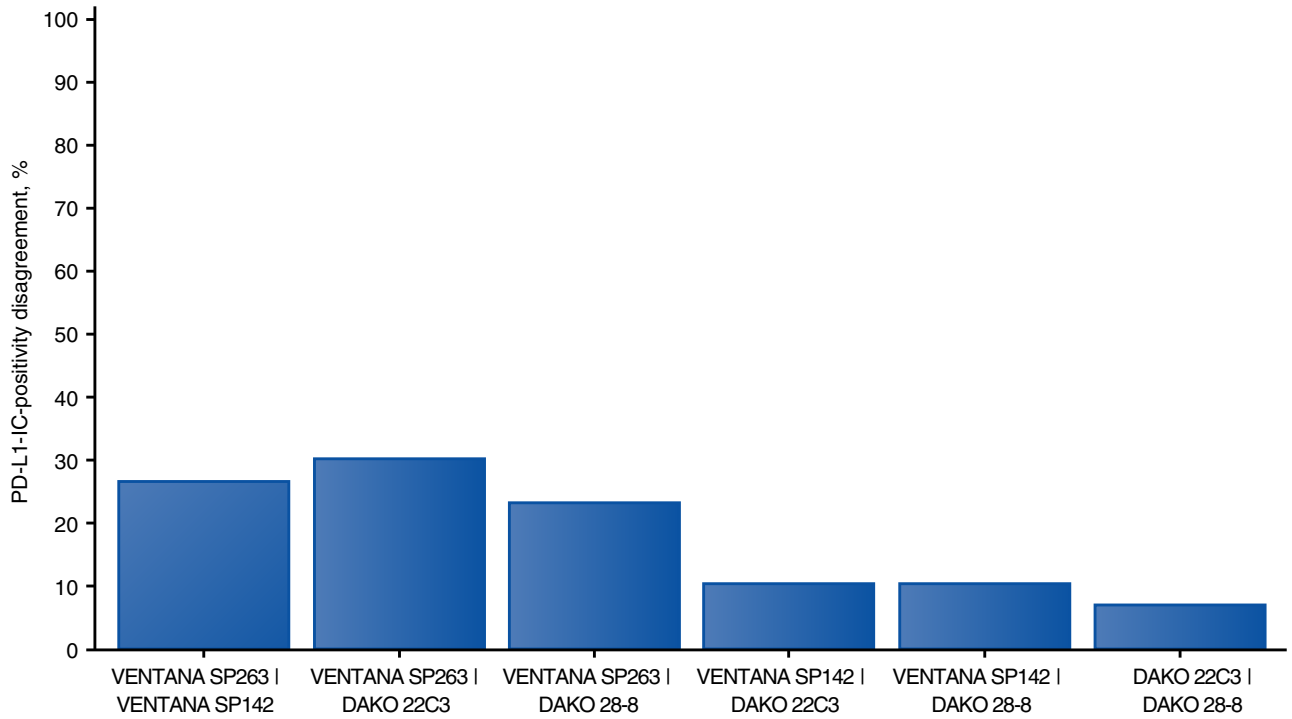


**Figure 3.** Heat-maps for PD-L1-IC-positivity agreement between assays (after averaging across the seven readers) for each sample (columns) when results were allocated to the retrospective  $\geq 1\%$  binary cut-off. Each column represents one sample and samples are ordered accordingly to their average PD-L1-IC-positivity among all readers and assays. For each specific assay, a sample is marked as green if PD-L1-IC-positivity averaged among the seven readers is  $\geq 1\%$  and red if PD-L1-IC-positivity is  $< 1\%$ . The number of samples with PD-L1-IC-positivity of  $\geq 1\%$  (green) and  $< 1\%$  (red) for each assay is given on the right. The bars for all assays summarise the number of assays for each sample that had PD-L1-IC-positivity of  $\geq 1\%$  (green) and  $< 1\%$  (red). IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1.

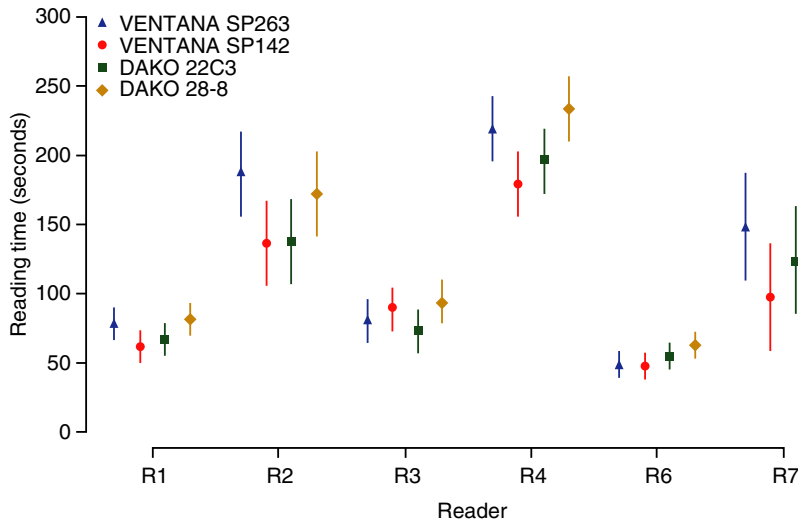
only 46% as PD-L1-IC-positive.<sup>34</sup> These discrepancies are not surprising, as scores were based on different criteria. It is critical, therefore, that PD-L1-IC-positivity scores are read correctly and taken into consideration for on-label treatment of patients with atezolizumab plus nab-paclitaxel.

Limitations of this study include screening and selection of cases based on SP142 only and the unavoidable inability to truly blind for assay information (stains were scored by experienced pathologists who may recognise the patterns). Based on the results presented, the distinct staining pattern of SP142 is likely not a key factor for discrepancies. Furthermore, as no corresponding data on patient outcomes are available, it is not clear whether SP263, with its different staining pattern, identifies patients with TNBC who may benefit from treatment with

atezolizumab plus nab-paclitaxel. In IMpassion130, 40.9% of patients were identified as having PD-L1-IC-positive tumours using SP142 at the cut-off of  $\geq 1\%$  of the tumour area, and a clear clinical benefit of anti-PD-L1 treatment was observed in this subgroup.<sup>12</sup> Assay comparisons in a subgroup of patients indicated that samples stained PD-L1-IC-positive with SP263 or 22C3, but simultaneously PD-L1-IC-negative with SP142, had limited clinical benefit.<sup>33</sup> Clinically, it is important that an assay identifies patients who respond well to a particular treatment rather than identifying a greater proportion of patients with PD-L1-IC-positive disease. The question of the most suitable assay for predicting benefit from cancer immunotherapy plus chemotherapy requires further clinical investigation, which is beyond the scope of this study.



**Figure 4.** Average PD-L1-IC-positivity agreement at the  $\geq 1\%$  cut-off between all assay pair combinations. IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1.



**Figure 5.** Reading times for PD-L1-IC-positivity using each of the four assays. IC, Tumour-infiltrating immune cell in the tumour area; PD-L1, Programmed death-ligand 1; R, Reader.

Our results indicate fair-to-strong analytical reproducibility and concordance of PD-L1-IC-positivity between SP142, 22C3 and 28-8, while SP263 was associated with higher PD-L1-IC-positivity, suggesting that SP142, 22C3 and 28-8 may be considered analytically interchangeable for testing PD-L1-IC-positivity in TNBC.

### Acknowledgements

We would like to thank Dr Nicole Krämer (Staburo GmbH, Munich, Germany) for help with statistical analyses, Dr Ulrich Roos (Roche Pharma AG, Grenzach-Wyhlen, Germany) for operational study management,



Sebastian Schefzyk (Roche Pharma AG) for critical revision of the manuscript and Drs Greg Harlow and Ehab ElGabry (both from Ventana Medical Systems, Inc., Tucson, AZ, USA) for their support with scoring training and for scoring of PD-L1-IC-positivity with VENTANA SP142. We would also like to thank Olga Seelbach, Ulrike Mühlthaler, Marion Mielke and Sarah Funkenhauser from the Comparative Experimental Pathology team at the Technical University of Munich and Anja Menges from the University Medical Center of the Johannes Gutenberg University Mainz, for their technical support. Support for third-party writing assistance for this manuscript, furnished by Katie Wilson PhD, of Health Interactions, was provided by Roche Pharma AG. Open access funding enabled and organized by Projekt DEAL.

## Conflict of interest

A.N. has received remuneration for travel, accommodation or expenses from Roche Pharma AG. J.U.A. was an employee of Roche Pharma AG at the time of the study, is currently an employee of F. Hoffmann-La Roche Ltd and holds stock in F. Hoffmann-La Roche Ltd. D.-C.W. has received funding from Roche Pharma AG. C.D. has received remuneration from Teva, Novartis, Pfizer, Roche Pharma AG and Amgen, has received consultancy or advisory fees from MSD Oncology, Amgen and Daiichi Sankyo and holds stock in Sividon Diagnostics/Myriad. A.L. has received remuneration for travel and accommodation expenses from Roche Pharma AG and Novartis and funding from Roche Pharma AG. P.S. has received remuneration from Roche Pharma AG and Nanostring and remuneration for travel, accommodation or expenses from Nanostring. H.-H.K. has received remuneration from Hannover Medical School, consultancy or advisory fees from Roche Pharma AG, Novartis, AstraZeneca and Genomic Health and funding from DFG and Deutsche Krebshilfe. U.S. has received personal fees, speaker honoraria and grants from Roche Pharma AG and MSD. G.B. has received remuneration for travel, accommodation or expenses from Roche Pharma AG and BMS, consultancy or advisory fees from Roche Pharma AG, BMS, MSD, AstraZeneca and Pfizer and funding from Roche Pharma AG, BMS and MSD. K.S. has received funding from Roche Pharma AG. M.K. 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 Diagnostisch GmbH and Busenfreundin GmbH and received funding from

Sphingotec, Deutsche Krebshilfe, DFG, the Senator Roesner Foundation and the Dr Pommer-Jung Foundation. S.H.-S. is an employee of Roche Pharma AG. M.F. is an employee of Ventana Medical Systems, Inc., which is part of the Roche Diagnostics Division, and holds stock in F. Hoffmann-La Roche Ltd. W.R. has received remuneration from Roche Pharma AG. W.W. has received consultancy or advisory fees from Roche Pharma AG, MSD, BMS, AstraZeneca, Astrella, Takeda, Lilly and Boehringer Ingelheim and funding from Roche Pharma AG, BMS, MSD and Bruker. All authors received support for third-party writing assistance for this manuscript, provided by Roche Pharma AG.

## Author contributions

J.U.A., G.B., S.H.-S. and W.W. were involved in the conception of the study. J.U.A., C.D., A.L., P.S., G.B., S.H.-S. and W.W. contributed to the study design. The first draft of the manuscript was written by A.N., J.U.A. and W.W. 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.

## Funding

This work was supported by Roche Pharma AG (Roche study ID: SL41336).

## Data Availability Statement

Qualified researchers may request access to individual patient-level data through the clinical study data request platform: <https://vivli.org/>. Further details on Roche's criteria for eligible studies are available from: <https://vivli.org/members/ourmembers/>. For further details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see: [https://www.roche.com/research\\_and\\_development/who\\_we\\_are\\_how\\_we\\_work/clinical\\_trials/our\\_commitment\\_to\\_data\\_sharing.htm](https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm). For readers of this publication, digital images of all samples stained with H&E, pan-cytokeratin or any of the four PD-L1 assays are available online: <http://www.roche.de/pd11testing>. The median PD-L1-IC score by readers is included in the case commentaries.

## References

1. Dent R, Hanna WM, Trudeau M, Rawlinson E, Sun P, Narod SA. Time to disease recurrence in basal-type breast cancers:

- effects of tumor size and lymph node status. *Cancer* 2009; **115**: 4917–4923.
2. Li X, Yang J, Peng L *et al.* Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Res. Treat.* 2017; **161**: 279–287.
  3. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines<sup>®</sup>): Breast Cancer, Version 2, 2020. Accessed 21 February 2020. Available at: [https://www.nccn.org/professionals/physician\\_gls/pdf/breast.pdf](https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf)
  4. Cardoso F, Senkus E, Costa A *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4). *Ann. Oncol.* 2018; **29**: 1634–1657.
  5. Wöckel A, Albert U-S, Janni W *et al.* The screening, diagnosis, treatment, and follow-up of breast cancer. *Dtsch. Arztebl. Int.* 2018; **115**: 316–323.
  6. Ditsch N, Untch M, Thill M *et al.* AGO recommendations for the diagnosis and treatment of patients with early breast cancer: update 2019. *Breast Care* 2019; **14**: 224–245.
  7. US Food and Drug Administration (FDA). FDA approves atezolizumab for PD-L1 positive unresectable locally advanced or metastatic triple-negative breast cancer. Accessed 21 February 2020. Available at: <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-atezolizumab-pd-1-positive-unresectable-locally-advanced-or-metastatic-triple-negative>
  8. Roche. European Commission approves Roche's Tecentriq in combination with Abraxane for people with PD-L1-positive, metastatic triple-negative breast cancer. Accessed 21 February 2020. Available at: <https://www.roche.com/investors/updates/inv-update-2019-08-29.htm>
  9. Denkert C, von Minckwitz G, Brase JC *et al.* Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J. Clin. Oncol.* 2015; **33**: 983–991.
  10. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 2013; **39**: 1–10.
  11. Schmid P, Adams S, Rugo HS *et al.* Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N. Engl. J. Med.* 2018; **379**: 2108–2121.
  12. Schmid P, Rugo HS, Adams S *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**: 44–59.
  13. Emens L, Loi S, Rugo H *et al.* Impassion130: efficacy in immune biomarker subgroups from the global, randomized, double-blind, placebo-controlled, phase III study of atezolizumab + nab-paclitaxel in patients with treatment-naïve, locally advanced or metastatic triple-negative breast cancer. *Cancer Res.* 2019; **79**: Abstract GS1-04.
  14. Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. *Immunol. Rev.* 2010; **236**: 219–242.
  15. Scheel AH, Dietel M, Heukamp LC *et al.* Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod. Pathol.* 2016; **29**: 1165–1172.
  16. Rimm DL, Han G, Taube JM *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**: 1051–1058.
  17. 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.* 2017; **12**: 208–222.
  18. Velcheti V, Patwardhan PD, Liu FX, Chen X, Cao X, Burke T. Real-world PD-L1 testing and distribution of PD-L1 tumor expression by immunohistochemistry assay type among patients with metastatic non-small cell lung cancer in the United States. *PLoS ONE* 2018; **13**: e0206370.
  19. Schwamborn K, Ammann JU, Knüchel R *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**: 599–608.
  20. 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**: 1059–1066.
  21. Tretiakova M, Fulton R, Kocherginsky M *et al.* Concordance study of PD-L1 expression in primary and metastatic bladder carcinomas: comparison of four commonly used antibodies and RNA expression. *Mod. Pathol.* 2018; **31**: 623–632.
  22. Eckstein M, Erben P, Kriegmair MC *et al.* Performance of the Food and Drug Administration/EMA-approved programmed cell death ligand-1 assays in urothelial carcinoma with emphasis on therapy stratification for first-line use of atezolizumab and pembrolizumab. *Eur. J. Cancer* 2019; **106**: 234–243.
  23. Zavalishina L, Tsimafeyeu I, Povilaitite P *et al.* RUSSCO-RSP comparative study of immunohistochemistry diagnostic assays for PD-L1 expression in urothelial bladder cancer. *Virchows Arch.* 2018; **473**: 719–724.
  24. Sommer U, Eckstein M, Ammann J *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; **15**: S1158–7673.
  25. Emens LA, Cruz C, Eder JP *et al.* Long-term clinical outcomes and biomarker analyses of atezolizumab therapy for patients with metastatic triple-negative breast cancer: a phase 1 study. *JAMA Oncol.* 2019; **5**: 74–82.
  26. Dennis E, Vennapusa B, ElGabry E *et al.* Robust and reproducible pathologist training for PD-L1 assessing tumour cells (TC) and immune cells (IC) utilising novel digital training platform. *Virchows Arch.* 2017; **471**: Abstract PS-21-004.
  27. Vennapusa B, Baker B, Kowanetz M *et al.* Development of a PD-L1 complementary diagnostic immunohistochemistry assay (SP142) for atezolizumab. *Appl. Immunohistochem. Mol. Morphol.* 2019; **27**: 92–100.
  28. Salgado R, Denkert C, Demaria S *et al.* The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol.* 2015; **26**: 259–271.
  29. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; **32**: 2847–2849.
  30. Noske A, Möbus V, Weber K *et al.* Relevance of tumour-infiltrating lymphocytes, PD-1 and PD-L1 in patients with high-risk, nodal-metastasised breast cancer of the German adjuvant intergroup node-positive study. *Eur. J. Cancer* 2019; **114**: 76–88.
  31. Sun WY, Lee YK, Koo JS. Expression of PD-L1 in triple-negative breast cancer based on different immunohistochemical antibodies. *J. Translat. Med.* 2016; **14**: 173.

32. 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; **30**: iii1–iii26.
33. Rugo H, Loi S, Adams S *et al.* Exploratory analytical harmonization of PD-L1 immunohistochemistry assays in advanced triple-negative breast cancer: a retrospective substudy of IMpassion130. *Cancer Res.* 2020; **80**: Abstract PD1-07.
34. Rugo H, Loi S, Adams S. Performance of PD-L1 immunohistochemistry (IHC) assays in unresectable locally advanced or metastatic triple-negative breast cancer (MTNBC): post-hoc analysis of IMpassion130. *Ann. Oncol.* 2019; **30**: v851–v934.
35. Casadevall D, Clavé S, Taus Á *et al.* Heterogeneity of tumor and immune cell PD-L1 expression and lymphocyte counts in surgical NSCLC samples. *Clin. Lung Cancer* 2017; **18**: 682–691.e685.
36. Scott ML, Scorer P, Lawson N *et al.* Assessment of heterogeneity of PD-L1 expression in NSCLC, HNSCC, and UC with Ventana SP263 assay. *J. Clin. Oncol.* 2017; **35**: e14502.
37. Rehman JA, Han G, Carvajal-Hausdorf DE *et al.* Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Mod. Pathol.* 2017; **30**: 340–349.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Immunohistochemistry.

**Data S2.** Calculation of the combined positivity score (CPS).

**Data S3.** Statistical methods.

**Data S4.** Inter-assay differences in programmed death-ligand 1 tumour cell (PD-L1-TC)-positivity.

**Table S1.** Antibody details.

**Table S2.** PD-L1-IC-positivity status per VENTANA SP142 and VENTANA SP263 in TNBC specimens in the screening cohort ( $n = 104$ ) and the study cohort ( $n = 30$ ).

**Table S3.** Clinicopathological characteristics of the study cohort ( $n = 30$ ).

**Table S4.** Inter-reader agreement Kappa values for each assay at the  $\geq 1\%$  cut-off for PD-L1-IC-positivity.

**Table S5.** Inter-assay agreement Kappa values for each reader at the  $\geq 1\%$  cut-off for PD-L1-IC-positivity.

**Table S6.** TIL-positive and -negative cases in the TNBC specimens from the screening ( $n = 104$ ) and study ( $n = 30$ ) cohorts.

**Table S7.** The association of TILs and PD-L1-IC status per VENTANA SP142 and VENTANA SP263 in the screening cohort.

**Table S8.** CPS ( $< 1$  or  $\geq 1$ ) per DAKO 22C3 for samples with PD-L1-IC-positivity of  $< 1\%$  and  $\geq 1\%$  per VENTANA SP142 at screening.

**Figure S1.** Differences in adjusted means of PD-L1-IC-positivity for each reader compared with the expert reader (R8).

**Figure S2.** Heat-maps for agreement on PD-L1-TC-positivity of  $\geq 1\%$  across readers for each of the samples (columns) for each of the assays.