

Article

Inter- and Intra-Observer Agreement of PD-L1 SP142 Scoring in Breast Carcinoma—A Large Multi-Institutional International Study

Mohamed Zaakouk ^{1,2} , Mieke Van Bockstal ^{3,4}, Christine Galant ^{3,4}, Grace Callagy ⁵, Elena Provenzano ^{6,7,8}, Roger Hunt ⁹, Corrado D'Arrigo ¹⁰, Nahla M. Badr ^{1,11}, Brendan O'Sullivan ¹², Jane Starczynski ¹³, Bruce Tanchel ¹³, Yasmeen Mir ¹⁴, Paul Lewis ¹⁵ and Abeer M. Shaaban ^{1,12,*} 

- ¹ Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, UK
- ² Cancer Pathology, National Cancer Institute, Cairo University, Cairo 12613, Egypt
- ³ Department of Pathology, Cliniques Universitaires Saint-Luc Bruxelles, 1200 Brussels, Belgium
- ⁴ Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, 1348 Brussels, Belgium
- ⁵ Discipline of Pathology, School of Medicine, Lambe Institute for Translational Research, University of Galway, H91 TK33 Galway, Ireland
- ⁶ NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, UK
- ⁷ Addenbrookes Hospital, Cambridge CB2 0QQ, UK
- ⁸ Department of Histopathology, Cambridge University NHS Foundation Trust, Cambridge CB2 0QQ, UK
- ⁹ Department of Histopathology, Wythenshawe Hospital, Manchester M23 9LT, UK
- ¹⁰ Poundbury Cancer Institute, Dorchester DT1 3BJ, UK
- ¹¹ Department of Pathology, Faculty of Medicine, Menoufia University, Shebin El-Kom 32952, Egypt
- ¹² Cellular Pathology, Queen Elizabeth Hospital Birmingham, Birmingham B15 2GW, UK
- ¹³ Cellular Pathology, Heart of England NHS Foundation Trust, Birmingham B9 5ST, UK
- ¹⁴ Pathology, Royal Liverpool and Broadgreen University Hospitals, Liverpool L7 8YE, UK
- ¹⁵ Medical School, Swansea University, Singleton Park, Swansea SA2 8PP, UK
- * Correspondence: a.shaaban@bham.ac.uk; Tel.: +44-121-371-3356



Citation: Zaakouk, M.; Van Bockstal, M.; Galant, C.; Callagy, G.; Provenzano, E.; Hunt, R.; D'Arrigo, C.; Badr, N.M.; O'Sullivan, B.; Starczynski, J.; et al. Inter- and Intra-Observer Agreement of PD-L1 SP142 Scoring in Breast Carcinoma—A Large Multi-Institutional International Study. *Cancers* **2023**, *15*, 1511. <https://doi.org/10.3390/cancers15051511>

Academic Editor: David Wong

Received: 1 November 2022

Revised: 15 February 2023

Accepted: 24 February 2023

Published: 28 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: PD-L1 analysis in TNBC is essential for selecting patients eligible for immunotherapy. Limited data are available on pathologists' concordance regarding PD-L1 assessment. Twelve pathologists of various expertise from three European countries digitally analysed 100 breast cancer core biopsies stained using the SP142 PD-L1 assay in two rounds. The overall inter-observer agreement among the pathologists was substantial. The intra-observer agreement was substantial to almost perfect. The expert scorers were more concordant in evaluating staining percentage compared with those of the non-experts. Challenging cases around the 1% cut-off value for positivity were identified and represented a small 6–8% proportion of all cases. The experts were more concordant in scoring those cases. The study shows reassuringly strong inter- and intra-observer concordance among pathologists in PD-L1 scoring. A proportion of low-expressors remain challenging to assess, and these would benefit from addressing the technical issues, testing a different sample and/or referring for expert opinions.

Abstract: The assessment of PD-L1 expression in TNBC is a prerequisite for selecting patients for immunotherapy. The accurate assessment of PD-L1 is pivotal, but the data suggest poor reproducibility. A total of 100 core biopsies were stained using the VENTANA Roche SP142 assay, scanned and scored by 12 pathologists. Absolute agreement, consensus scoring, Cohen's Kappa and intraclass correlation coefficient (ICC) were assessed. A second scoring round after a washout period to assess intra-observer agreement was carried out. Absolute agreement occurred in 52% and 60% of cases in the first and second round, respectively. Overall agreement was substantial (Kappa 0.654–0.655) and higher for expert pathologists, particularly on scoring TNBC (6.00 vs. 0.568 in the second round). The intra-observer agreement was substantial to almost perfect (Kappa: 0.667–0.956), regardless of PD-L1 scoring experience. The expert scorers were more concordant in evaluating staining percentage compared with the non-experienced scorers ($R^2 = 0.920$ vs. 0.890). Discordance predominantly occurred in low-expressing cases around the 1% value. Some technical reasons contributed to the discordance. The study shows reassuringly strong inter- and intra-observer concordance among

pathologists in PD-L1 scoring. A proportion of low-expressors remain challenging to assess, and these would benefit from addressing the technical issues, testing a different sample and/or referring for expert opinions.

Keywords: PD-L1; breast cancer; VENTANA SP142; triple-negative

1. Introduction

Advances in biomarker assessment, companion diagnostics and genomics have revolutionised the way breast cancer is currently classified and managed [1,2]. The immune microenvironment of solid tumours, including in breast cancer, plays a pivotal role in tumour development and progression [3–5]. Cancer cells can evade the regulatory pathways of Programmed death-1 (PD-1) and its ligand (PD-L1), thus overcoming the cytotoxic effect of T cells. Immune checkpoint blockades using anti-PD-L1 inhibitors have been investigated in various trials in lung, melanoma and, more recently, breast cancer, with confirmed efficacy [6–8]. This has led to the approval of immune modulators for the treatment of PD-L1-positive breast cancer, and this is currently being incorporated in various guidelines [9]. The first-approved and most established immune checkpoint inhibitor in breast cancer is atezolizumab, for which a companion diagnostic assay (the VENTANA SP142) is required for selecting patients eligible for this drug.

The limited data available in the literature on non-breast cancer suggest the poor reproducibility of PD-L1 SP142 scoring [10]. Some studies compared the performance of various PD-L1 assays [11,12], and only few analysed pathologist concordance in the scoring of breast cancer [13–15]. Those latter studies were small and heterogeneous, with some including training sets [14]. Furthermore, the nature of discordant cases was not analysed, nor was there an assessment of the intra-observer agreement or the effect of the pathologist's experience. In addition, all previous studies focused on TNBC, and, therefore, information on pathologist concordance in the scoring of PD-L1 in HER2-positive and/or luminal breast cancer does not exist. Emerging data suggest cross-talk between HER2 and PD-L1 and potentially support the use of immunotherapy in HER2-positive breast cancer [16]. PD-L1 expression is correlated with the response to neoadjuvant chemotherapy in HER2-positive breast cancer [17].

We therefore aimed to assess the inter- and intra-observer concordance of breast pathologists of various expertise and geographical locations in reporting a large cohort of PD-L1 SP142-stained invasive breast carcinomas of various molecular subtypes to assess if particular molecular subtypes would be more or less prone to poor inter-observer concordance. We also sought to analyse discordant cases in detail to gain insight into the reasons for discrepancies in PD-L1 results, allowing for a subsequent search for strategies on how to tackle them.

2. Materials and Methods

Core biopsies from a total of 100 cases of primary breast cancers were included in the study. Cases were selected retrospectively from the files of a single large UK institution (Queen Elizabeth Hospital Birmingham) to include all molecular subtypes with enrichment for the TNBC group.

First, 4 µm sections of formalin-fixed, paraffin-embedded tumour blocks were cut and stained using the VENTANA SP142 anti-PD-L1 rabbit monoclonal primary antibody and a VENTANA Benchmark ULTRA automated staining platform, according to the manufacturer's protocol. A section from a cell block containing three cell lines with various staining intensities and a section of normal tonsil were included as on-slide controls. Paired H&E sections and PD-L1-stained immunohistochemistry slides were digitally scanned using a Leica Aperio AT2 slide scanner at x40 and uploaded to the University of Birmingham digital platform via a secure link: <https://eslidepath.bham.ac.uk>, Last accessed 23 February

2023. Each participant was provided with a unique username and password to allow for access to the digital platform for whole slide scoring. Twelve pathologists from eight institutions representing three European countries (United Kingdom, Republic of Ireland, Belgium) evaluated all cases in round one, of whom 10 re-scored the same cases in round two, separated by at least 3 months of a washout period designed to assess intra-observer variability. All pathologists had previously received Roche training for SP142 PD-L1 scoring in TNBC and passed a proficiency test.

PD-L1 SP142 Immune cell (IC) scoring was conducted according to the recommended scoring algorithm [18], using a cut-off value of $\geq 1\%$ to indicate positivity. In addition, the pathologists were asked to provide their percentage of immune cells with positive staining for each case, including those cases scored as negative. All scorers completed a survey assessing their experience in breast pathology reporting as well as their training and real-life reporting of PD-L1.

Statistical Analysis

The data were tabulated and statistically analysed using the SPSS (IBMS) software version 28. We used standard statistical analyses for assessing intra/inter-rater concordance/agreement, which have been previously described [19]. Intraclass correlation coefficient (ICC), which is a measure of the reliability of ratings (using median percentage scores), was used to determine if subjects/items can be rated reliably by different raters. ICC is a descriptive statistic used to assess the consistency or reproducibility of quantitative measurements made by different observers measuring the same quantity. The value of an ICC can range from 0 to 1, with 0 indicating no reliability among raters and 1 indicating perfect reliability among raters. The ICC results are interpreted as follows: values < 0.5 indicate poor reliability, values from 0.5 to 0.75 indicate moderate reliability, values from 0.75 to 0.9 indicate good reliability and values greater than 0.9 indicate excellent reliability [20]. In our study, we used a Two-Way Random model, testing both the consistency and the absolute agreement relationships and the mean of ratings as the unit of measurement.

Fleiss multiple-rater Kappa statistics of inter-observer and intra-observer agreement for designating cases as PD-L1-positive versus -negative using a cut-off value of 1% were calculated. Fleiss' Kappa κ is a measure of inter-rater agreement used to determine the level of agreement between two or more raters when the method of assessment, known as the response variable, is measured on a categorical scale. The Kappa results are interpreted as follows: values ≤ 0 indicate no agreement, values from 0.01 to 0.20 indicate none to slight agreement, values from 0.21 to 0.40 indicate fair agreement, values from 0.41 to 0.60 indicate moderate agreement, values from 0.61 to 0.80 indicate substantial agreement and values from 0.81 to 1.00 indicate almost perfect agreement.

A case was regarded as PD-L1 positive or -negative if more than 50% of the participants designated it as positive or negative, respectively. The consensus score was considered a majority score if 67% or more of the participants agreed on the categorisation. If all participants agreed (100%), this was regarded as absolute agreement (AA). The cases with agreement less than 67% and above 50% were considered challenging. In cases of no agreement (50% or less), a case was considered as PD-L1-positive or -negative based on the consensus of the experienced pathologists only.

Scatter plots were used to visualise percentage PD-L1 scores, and the strength of the relationship between scores was expressed as a squared correlation coefficient (R^2). All analyses were supervised by an expert in pathology informatics (PL).

An outline of the study methodology is shown in Figure 1.

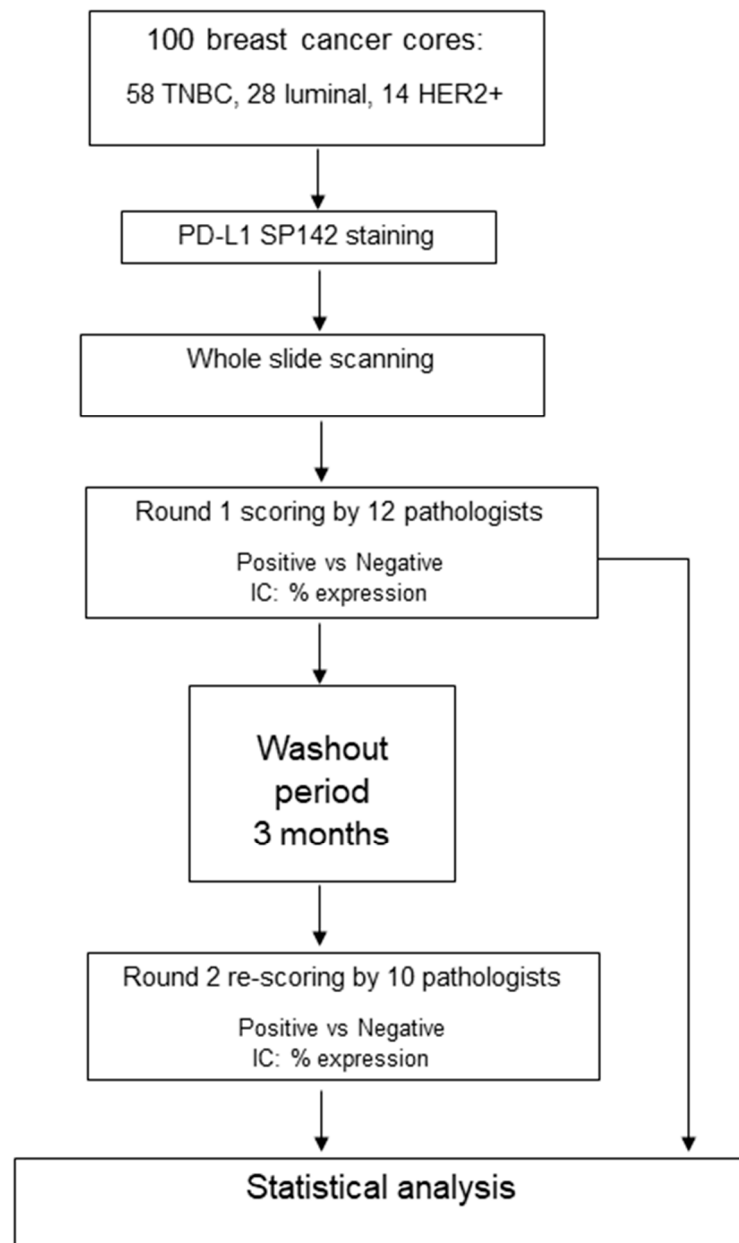


Figure 1. A flowchart showing the outline of the study.

3. Results

3.1. Cohort Characteristics

A total of 100 breast cancers were assessed, comprising 29/93 (33.3%) grade 2 and 62/93 (66%) grade 3 cases, while 7 cases had missing grades. The patient ages ranged from 42 to 59 years, with a median of 49 years. Fifty-eight carcinomas were triple-negative, 28 were luminal and 14 were Her2-positive. All cases were evaluated independently by twelve pathologists, including nine specialist consultant breast pathologists, of whom six had 1–3 years' experience in PD-L1 scoring, as per the survey responses. All scorers had significant experience in breast pathology reporting, and six had experience in scoring PD-L1 SP142 in TNBC in routine practice. A consultant Biomedical Scientist and two trainee pathologists were among the scorers. Ten pathologists scored round two, following a washout period. The overall percentage of PD-L1 positivity for all of the breast cancer molecular subtypes was 36–38%, and the highest was for TNBC (55%) (Table 1).

Table 1. Frequency of PD-L1 positivity in breast cancer in both rounds, stratified by the molecular type.

	No.	First Round		Second Round	
		Positive (%)	Negative (%)	Positive (%)	Negative (%)
TNBC	58	32 (55%)	26 (45%)	32 (55%)	26 (45%)
Median (range)		4 (0.75–30)	0 (0–1)	5 (0.5–30)	0 (0–1)
Luminal	28	4 (14%)	24 (86%)	4 (14%)	24 (86%)
Median (range)		2 (1–4)	0 (0–0.75)	3.5 (1.5–5)	0 (0–0.5)
Her2-positive	14	2 (14%)	12 (86%)	1 (7%)	13 (93%)
Median (range)		5.5 (1–10)	0 (0–0.5)	10	0 (0–0.5)
Total	100	38 (38%)	62 (62%)	36 (36%)	64 (64%)
Median (range)		2 (0.75–30)	0 (0–1)	5 (0.5–30)	0 (0–1)

3.2. Inter-Observer Agreement and Pathologist Experience

The Kappa of the inter-observer agreement between the participants in classifying cases as PD-L1-positive vs. -negative and the number of cases with absolute agreement (AA) for rounds one and two were calculated. The overall agreement was substantial (Kappa 0.654 and 0.655 for the first and second rounds, respectively) (Table 2).

Table 2. Absolute agreement in scoring among raters in the two rounds.

	Raters	P1	P2	P3	P4 ^c	P5 ^c	P6 ^c	P7 ^e	P8 ^e	P9 ^e	P10 ^e	P11 ^e	P12 ^e
First Round	Neg	62	64	61	58	63	68	75	51	64	67	63	63
	Pos	38	35	31	41	37	32	25	49	36	33	37	34
	Total	100	99	92	99	100	100	100	100	100	100	100	97
	Kappa	0.654											
	AA	52/100 cases; 36 scored negative and 16 scored positive											
Second Round	Neg	60	64		64	46	62		52	72	69	69	66
	Pos	40	35		35	54	37		48	28	31	31	30
	Total	100	99		99	100	100		100	100	100	100	97
	Kappa	0.655											
	AA	60/100 cases; 40 scored negative and 20 scored positive											

^e Experienced consultant, ^c Consultant, (AA) Absolute agreement. P3 and P7 did not score the second round.

There was absolute agreement on scoring cases as either positive or negative in 52 cases in the first round (Figure 2A–D). This increased to 60 cases in the second round (Table 3). A further 42 and 32 cases achieved majority agreement in the first and second rounds, respectively. Overall, the Kappa value, for all cases, was similar between experienced pathologists and those without considerable experience in PD-L1 reporting. However, it was higher for expert pathologists scoring PD-L1 in the TNBC group (0.6 vs. 0.568 in the second round) (Table 4).

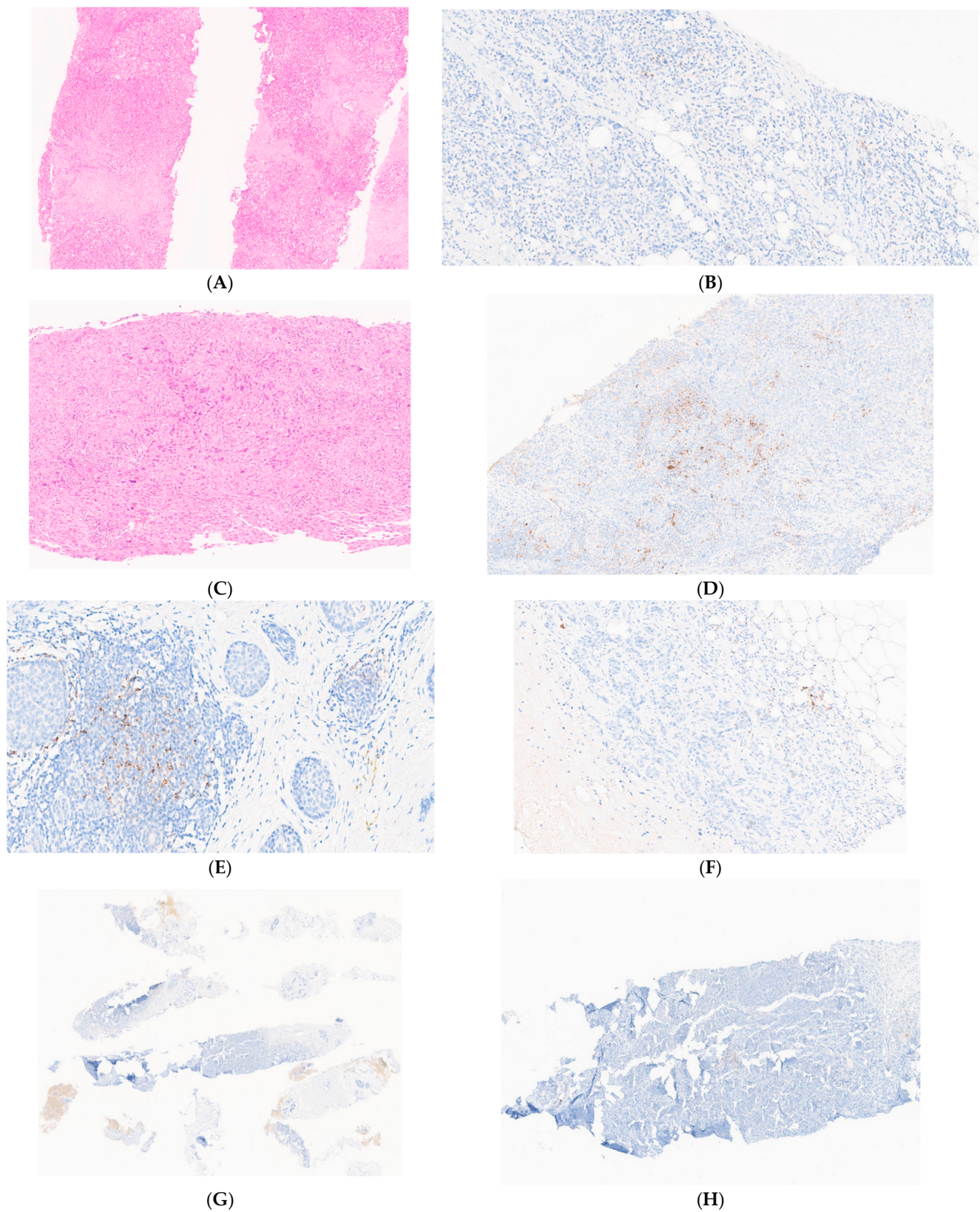


Figure 2. Examples of breast cancer PD-L1 scores using the Ventana SP142 assay and challenging cases: (A) H&E staining of three cores of invasive no-special-type carcinoma ($\times 50$); (B) Only focal PD-L1 staining is noted ($<1\%$), and the case was classified as PD-L- negative ($\times 100$); (C) H&E staining

of one core of invasive no-special-type carcinoma ($\times 100$); (D) Higher magnification of PD-L1 immunohistochemistry shows strong positivity with absolute agreement among all scorers in both rounds ($\times 100$); (E) Challenging case due to uncertainty as to whether the PD-L1 staining observed is associated with in situ or invasive carcinoma ($\times 100$); (F) Challenging cases showing low levels of expression. Experts' consensus was to designate the case as PD-L1-negative ($\times 100$); (G) A case with no consensus in either round. Low-power view showing areas of tumour necrosis and background staining ($\times 15$); (H) Higher magnification showing focal expression in tumour stroma and adjacent to an area of necrosis ($\times 50$). This case was also challenging for experts; in the first round, it showed low agreement (60% as negative), and in the second round, it showed no agreement (50%).

Table 3. Agreement categories in the first and second scoring rounds.

Round		Consensus (Agreement)			No Agreement
		Majority		Challenging/Low Agreement	$\leq 50\%$
		100% (AA)	67–99%	<67–>50%	
First	Negative	36	24	2	0
	Positive	16	18	4	0
		52	42	6	0
	Total	94		6	0
			100		0
Second	Negative	40	20	2	2
	Positive	20	12	4	0
		60	32	6	2
	Total	92		6	2
			98		2

Table 4. Fleiss Kappa of agreement between the pathologists in both rounds.

	Fleiss Kappa First Round			Fleiss Kappa Second Round		
	Scoring Categories			Scoring Categories		
	Overall (TNBC)	NEG	POS	Overall (TNBC)	NEG	POS
All	0.654 (0.61)	0.660	0.678	0.655 (0.602)	0.656	0.669
Consultants	0.663 (0.616)	0.664	0.673	0.633 (0.568)	0.636	0.650
Experienced	0.659 (0.642)	0.661	0.672	0.674 (0.600)	0.677	0.695

3.3. Concordance of PD-L1 Percentage Expression

When the median percentage of PD-L1 expression was considered, the expert pathologists had a higher and tighter concordance compared with the non-experts ($R^2 = 0.920$ vs. 0.89). The overall concordance was excellent ($R^2 = 0.935$). The distribution of the percentage scoring among all raters (those with and without experience in PD-L1 routine reporting) in both rounds is shown in Figure 3A–E.

3.4. Reasons for Discordance

Ten cases were regarded as challenging, with low (<67–>50%) or no agreement (<50%), most of which (8/10) were of the TNBC phenotype (Table 5). All ten cases had a low PD-L1 score, with a median range of 0.5–1%, highlighting the difficulties in classifying cases close to the cut-off value of 1%. Four cases were challenging in both rounds, indicating of the innate difficulty of the cases, regardless of the pathologists' expertise in PD-L1 scoring. We analysed the reasons for the difficulties in scoring those cases by reviewing the digital

images and referring to pathologists' comments on scoring. Those cases were reassuringly recognised as difficult by most scorers due to the nature of the tumour and/or technical issues. Reasons for discordance included uncertainty as to the presence/extent of the in situ carcinoma, a small amount of invasive carcinoma, positive staining around the normal mammary epithelium, very focal staining, background staining and staining within areas of necrosis (Figure 2E–H).

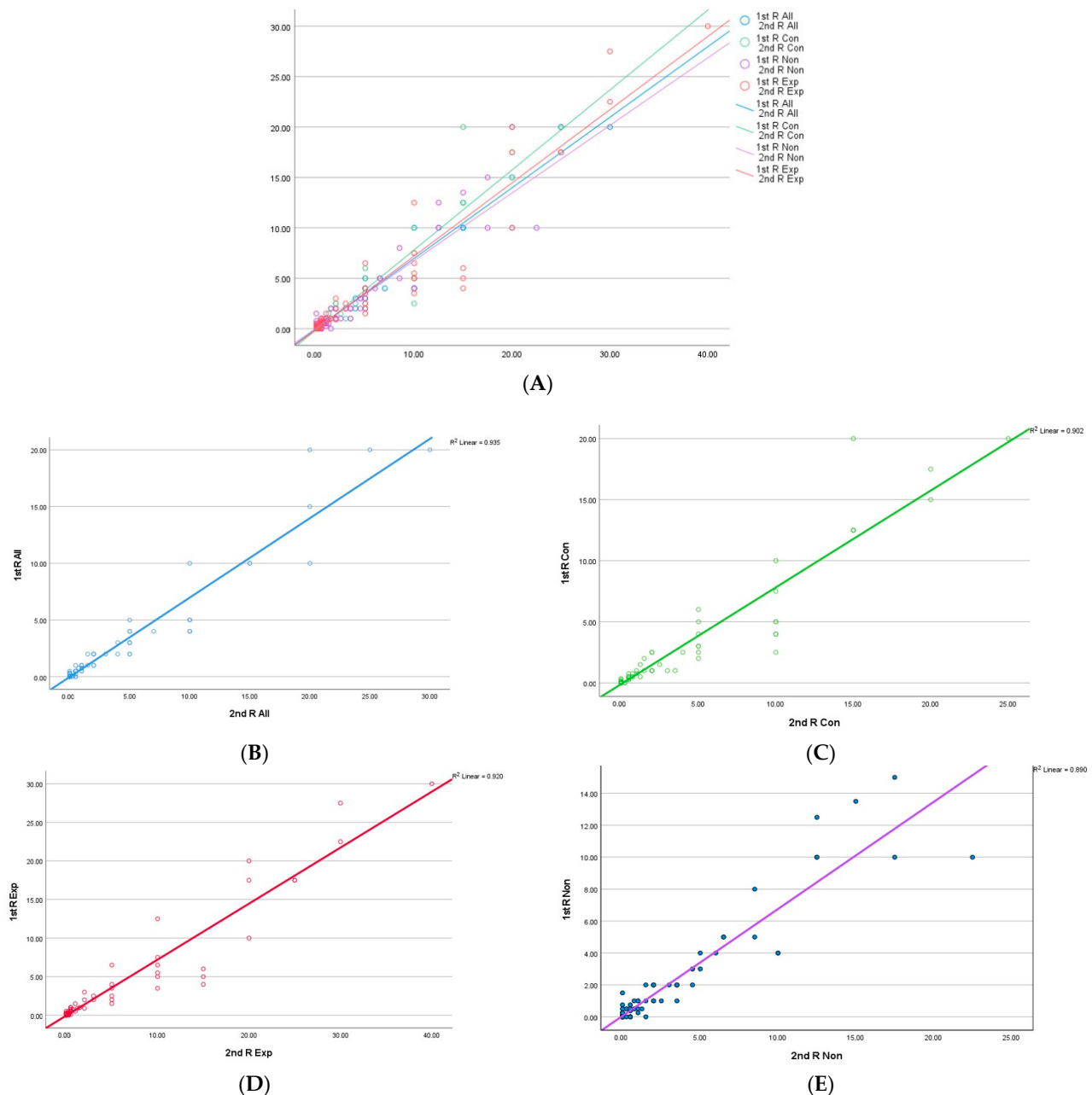


Figure 3. Scatter plot showing the distribution of the median percentage of PD-L1 in the four groups; experienced consultants (Exp; red line), non-experienced consultants (Con; green line) and all (All; blue line) participants. (A) The distribution of percentage scores among all scorers in both rounds; (B) All pathologists' (including non-experts and trainees) percentage scores in both rounds ($R^2 = 0.935$); (C) All consultants' (including experts and non-experts) percentage scores in both rounds ($R^2 = 0.902$); (D) Experienced consultants' percentage scores in both rounds ($R^2 = 0.920$); (E) Non-experienced pathologists' (including trainees) percentage scores in both rounds ($R^2 = 0.89$).

Table 5. Distribution of ten challenging (low concordance and no agreement) cases in both rounds.

Type	FIRST ROUND									SECOND ROUND								
	ALL (12)	PD-L1 Status	M	NON (6)	PD-L1 Status	M	EXP (6)	PD-L1 Status	M	All (10)	PD-L1 Status	M	Non (5)	PD-L1 Status	M	EXP (5)	PD-L1 Status	M
TNBC	6/11 (55%)	—	0.5	3/6 (50%)	—	0.75	3/5 (60%)	—	0.5	6/9 (67%)	—	0.5	4/5 (80%)	—	0	2/4 (50%)	—	1
Her2	7/12 (58%)	—	1	4/6 (67%)	—	1	3/6 (50%)	—	0.75	7/10 (70%)	—	0.5	3/5 (60%)	+	1	5/5 (100%)	—	0.5
TNBC	7/12 (58%)	—	0.5	4/6 (67%)	—	0.25	3/6 (50%)	—	0.75	6/10 (60%)	+	1	4/5 (80%)	+	1	3/5 (60%)	—	0.5
TNBC	7/12 (58%)	—	1	3/6 (50%)	—	1	4/6 (67%)	—	1	6/10 (60%)	—	1	3/5 (60%)	+	1	4/5 (80%)	—	0.75
TNBC	7/12 (58%)	+	1	4/6 (67%)	+	1	3/6 (50%)	—	0.75	5/9 (56%)	+	1	2/4 (50%)	—	2	3/5 (60%)	+	0.75
TNBC	7/12 (58%)	—	1	4/6 (67%)	+	1	5/6 (83%)	—	0.5	5/10 (50%)	—	1	4/5 (80%)	+	1.5	4/5 (80%)	—	0.75
TNBC	9/12 (75%)	+	1	4/6 (67%)	+	1	5/6 (83%)	+	1	6/10 (60%)	+	2	4/5 (80%)	+	2.5	3/5 (60%)	—	0.5
TNBC	8/12 (67%)	+	1	3/6 (50%)	—	0.5	5/6 (83%)	+	1	6/10 (60%)	—	1	3/5 (60%)	—	0.5	4/5 (80%)	+	1.5
TNBC	11/12 (83%)	—	0.5	6/6 (100%)	—	0.5	5/6 (83%)	—	0.5	6/10 (60%)	—	0.5	3/5 (60%)	+	1	4/5 (80%)	—	0.5
Lum	8/12 (75%)	+	1	4/6 (67%)	+	1	4/6 (67%)	+	1	5/10 (50%)	—	1.5	4/5 (80%)	+	3.5	4/5 (80%)	—	0.5
AGREEMENT	No	0			3/10; 30%			3/10; 30%			2/10; 20%			1/10; 10%			1/10; 10%	
	Low	6/10; 60%			0			1/10; 10%			6/10; 60%			4/10; 40%			3/10; 30%	
	High	4/10; 40%			7/10; 10%			6/10; 60%			2/10; 20%			5/10; 50%			6/10; 60%	

PD-L1 status: negative (—) or positive (+); (M) Median percentage score; (ALL) refers to all scorers, which were then divided into (NON) Non-expert participants in PD-L1 scoring and (EXP) Expert pathologists. Cases in green represent challenging cases in the first round, those in orange represent challenging ones in both rounds and those in grey represent challenging and no-agreement cases in the second round only. The yellow cases represent contradictory agreement of non-experts in relation to the experts' agreement, and the blue cases represent cases with no agreement (50%). Agreement levels are categorised into no-agreement (50%), low-agreement (>50%–<67%) and high-agreement (>67%).

It is of note that the consensus in scoring those challenging cases among the expert pathologists ranged from no agreement (50%) to absolute agreement (100%), with a higher percentage of the former in the first round (3/10; 30%) compared to the second (1/10; 10%). For the experts, concordance improved in the second round, with all but one case showing absolute agreement. Strong to almost perfect agreement among the experts was seen in 6/10 of those challenging cases, in both rounds. For the non-experts, the proportion of cases with low or no agreement was higher than that for the experts and increased from 30% in the first round to 50% in the second round (Table 5).

3.5. Inter- and Intra-Observer Agreement

Cohen's Kappa was calculated to assess the inter-observer agreement between each of the scoring pathologists and the intra-observer agreement for each scorer across the two rounds. Inter-observer agreement was, overall, moderate (0.5) to substantial (0.75) (Table 6). The highest Kappa values for inter-observer agreement were 0.871 (in the first round) and 0.88 (in the second round), while the lowest values were moderate: 0.475 (in the first round) and 0.498 (in the second round). The intra-observer agreement was substantial to almost perfect, ranging from 0.667 to 0.956 (Table 6).

Table 6. Inter- and intra-observer agreement for all scoring pathologists.

	Consensus 1	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Consensus 2	0.912	0.851	0.78		0.823	0.629	0.715		0.737	0.747	0.819	0.865	0.884
P1	<i>0.892</i>	0.832	0.766		0.679	0.724	0.787		0.758	0.649	0.719	0.762	0.733
P2	<i>0.765</i>	<i>0.747</i>	0.722		0.735	0.551	0.695		0.575	0.562	0.729	0.774	0.745
P3	<i>0.786</i>	<i>0.768</i>	<i>0.607</i>										
P4	<i>0.823</i>	<i>0.762</i>	<i>0.654</i>	<i>0.641</i>	0.956	0.587	0.669		0.631	0.606	0.637	0.728	0.699
P5	<i>0.74</i>	<i>0.723</i>	<i>0.64</i>	<i>0.696</i>	<i>0.607</i>	0.667	0.682		0.801	0.498	0.554	0.515	0.539
P6	<i>0.798</i>	<i>0.737</i>	<i>0.741</i>	<i>0.632</i>	<i>0.617</i>	<i>0.713</i>	0.732		0.632	0.635	0.688	0.643	0.656
P7	<i>0.718</i>	<i>0.659</i>	<i>0.579</i>	<i>0.717</i>	<i>0.669</i>	<i>0.54</i>	<i>0.634</i>						
P8	<i>0.678</i>	<i>0.658</i>	<i>0.533</i>	<i>0.543</i>	<i>0.613</i>	<i>0.678</i>	<i>0.577</i>	<i>0.475</i>	0.94	0.552	0.574	0.614	0.661
P9	<i>0.891</i>	0.871	<i>0.745</i>	<i>0.764</i>	<i>0.715</i>	<i>0.72</i>	<i>0.733</i>	<i>0.744</i>	<i>0.658</i>	0.772	0.784	0.64	0.826
P10	<i>0.822</i>	<i>0.76</i>	<i>0.634</i>	<i>0.831</i>	<i>0.687</i>	<i>0.649</i>	<i>0.704</i>	<i>0.711</i>	<i>0.597</i>	<i>0.801</i>	0.862	0.762	0.88
P11	<i>0.87</i>	<i>0.808</i>	<i>0.724</i>	<i>0.649</i>	<i>0.738</i>	<i>0.743</i>	<i>0.801</i>	<i>0.586</i>	<i>0.638</i>	<i>0.763</i>	<i>0.737</i>	0.778	0.784
P12	<i>0.843</i>	<i>0.735</i>	<i>0.646</i>	<i>0.758</i>	<i>0.708</i>	<i>0.643</i>	<i>0.7</i>	<i>0.687</i>	<i>0.563</i>	<i>0.732</i>	<i>0.749</i>	<i>0.732</i>	0.906

Figures in italics below the equatorial bordered cells represent the values of the first round, while those in bold represent the second round. The equatorial bordered cells (in bold red font) represent the intra-observer agreement for each participant, as scored in both rounds. Cell shading colours reflect the level of agreement as follows; light green for almost perfect agreement (0.81–1), orange for substantial agreement (0.61–0.8) and light red for moderate agreement (0.41–0.6).

3.6. Intraclass Correlation Coefficient (ICC)

ICC was used to assess the reliability of scoring between different groups of raters using the median percentage expression. The ICC for different groups (all scorers, experienced scorers and non-experienced scorers) ranged from moderate (0.5–0.75) to excellent (>0.9), with the predominance of the latter (Table 7). The highest ICC was 0.974 (between all scorers in first round and experienced ones in the second round), while the lowest value was 0.619 (between the non-experienced in the first round and the experienced in the second round).

Table 7. Intraclass correlation coefficient for all groups of pathologists.

	ALL-1	EXP-1	NON-1	ALL-2	EXP-2	NON-2
ALL-1		0.907	0.931	0.906	0.768	0.932
EXP-1	0.915		0.772	0.974	0.913	0.919
NON-1	0.933	0.788		0.781	0.619	0.876
ALL-2	0.919	0.974	0.804		0.911	0.946
EXP-2	0.798	0.923	0.655	0.919		0.792
NON-2	0.936	0.920	0.891	0.949	0.808	

Figures in bold/italics below the equatorial grey cells represent values of ICC calculated according to the consistency of assessment, while values below equatorial cells represent ICC calculated according to absolute agreement. (ALL): All scorers' median percentage; (EXP): Experienced scorers' median percentage; (NON): Non-experienced scorers' median percentage. Cells' shading colours reflect the level of reliability as follows: Red for moderate reliability (0.5–0.75); Orange for good reliability (0.75–0.9); Green for excellent reliability (greater than 0.9).

3.7. Intra-Observer Agreement and Scoring Reliability in Relation to Pathologists' Experience

All scorers had significant experience in breast pathology reporting, but only six scored PD-L1 SP142 in breast cancer in routine practice. The experience in PD-L1 reporting did not appear to affect the intra-observer agreement, with all scorers showing substantial or almost perfect agreement. On the other hand, the intra-observer reliability in the percentage assessment of PD-L1 expression was higher for experienced pathologists compared with non-experienced pathologists (Table 8).

Table 8. Intra-observer concordance based on pathologists' experience in scoring PD-L1.

Rater	Position	Experience as a Breast Reporting Pathologist (years)	Experience in SP142 PD-L1 Reporting (years)	Previous Training in SP142 PD-L1 Reporting (Provider)	Intra-Observer Agreement (Cohen's Kappa/Level of Agreement)	Intra-Observer Reliability (ICC/Level of Reliability)
P1	Trainee Pathologist	12	0	Roche	0.832/Almost perfect	0.826/Good
P2		12	0	Roche	0.722/Substantial	0.525/Moderate
P3	Consultant Scientist	N/A	0	Roche	N/A/N/A	N/A/N/A
P4	Consultant Pathologist	20	0	N/S	0.956/Almost perfect	0.852/Good
P5		21	0	Roche	0.667/Substantial	N/A/N/A
P6		25	0	None	0.732/Substantial	0.770/Good
P7		25	3	Roche	N/A/N/A	N/A/N/A
P8		29	1	Roche	0.94/Almost perfect	0.935/Excellent
P9		10	2	Roche	0.772/Substantial	0.933/Excellent
P10		25	2	Roche	0.862/Almost perfect	0.920/Excellent
P11		30	3	Local	0.778/Substantial	0.756/Good
P12		22	2	Roche	0.906/Almost perfect	0.929/Excellent

(N/S) Not stated; (N/A) Not applicable.

4. Discussion

We present comprehensive data of a large PD-L1 concordance cohort, scored twice by pathologists from eight institutions, representing three countries. Our data show reassuring inter- and intra-observer agreements, which were the highest among experts, and highlight cancers with low levels of PD-L1 expression as the most challenging in classifying as either PD-L1-positive or -negative.

Unlike standard diagnostic and prognostic markers for breast cancer, SP142 PD-L1 immunohistochemistry is assessed in the immune micro-environment of breast cancer and not in the neoplastic cells themselves. PD-L1 expression in foci of ductal carcinoma in situ (DCIS), necrotic debris, normal mammary tissue and normal nodal tissue is excluded. Therefore, experience in both tumour morphology and PD-L1 assessment is required and may affect the reproducibility of scoring.

Few studies, summarised in Table 9, have addressed the consistency of PD-L1 reporting among pathologists. A prospective multi-institutional study showed the poor reproducibility of PD-L1 scoring, with pathologists disagreeing on the classification of cases as PD-L1-positive or -negative in over half of the scored cases, and the with complete agreement of SP-142 scoring in only 38% of cases [21]. In a cohort of 426 tumours of Chinese women, the concordance between two pathologists in PDL-1 scoring was 78.2%, with a Kappa value of 0.567, and 61.4% in primary tumours and nodal metastasis, respectively, indicating moderate agreement [22]. Using “Observers Needed to Evaluate Subjective Tests” (ONEST), Reisenbichler et al. [21] reported a decreased overall percentage agreement with the increase in the number of pathologists assessing each case, with the lowest concordance at eight pathologists or more. Another study of 79 PD-L1 SP142-stained breast cancers scored by experienced breast pathologists at the Memorial Sloan-Kettering Cancer Centre revealed strong agreement [23]. Our data, based on a larger cohort of TNBC cases, confirm the substantial agreement and show that concordance was higher among experts than among those with no experience in reporting PD-L1. More importantly, the agreement among experts was observed as substantial to perfect in those challenging cases, and those experts showed a much higher consistency in reporting challenging, low-expressing TNBC, a finding that is relevant to clinical practice. This is in accordance with findings in other biomarkers [24] and reflects the importance of testing at regional institutions with quality-assured protocols and experienced scorers and the value of discussing/referring difficult/equivocal cases to expert pathologists for their opinions.

While several antibodies/assays for PD-L1 assessment are available (e.g., 22C3, 28-8, SP142, SP263 and 73-10), the VENTANA Roche SP142 assay is the only companion FDA- and CE-IVD (European Commission in vitro diagnostics)-approved test for atezolizumab therapy. An expert round table in 2019 [25] recommended the assay as the only approved companion diagnostic for selecting patients for immunotherapy and recommended using the primary tumour samples, where available, over metastases for assessment. In the UK, atezolizumab plus chemotherapy, and its companion diagnostic assay, were granted approval by the National Institute of Health and Care Excellence (NICE) for the treatment of locally advanced/metastatic PD-L1-positive TNBC. More recently, pembrolizumab plus chemotherapy has been approved for the same indication for PD-L1-positive TNBC using the companion diagnostic Agilent 22C3 assay.

In this study, we assessed both the inter- and intra-observer concordance among the participating pathologists. It is notable that the intra-observer concordance was high (0.667 to 0.956) among both expert and non-expert pathologists in PD-L1 scoring, indicating that pathologists are likely to stick to their parameters on scoring. When the median percentage of PD-1 expression was compared among the raters, the highest ICC (0.974) was achieved among experienced raters in the second round. We observed the lowest concordance value of 0.619 when comparing non-experienced to experienced scorers. Similarly, a higher concordance among those experienced in PD-L1 scoring (93.3%) compared with non-experts (81.5%) was previously reported by Pang et al. [26].

While, overall, there was a high concordance among pathologists in PD-L1 SP142 scoring, some cases were challenging to score. Those cases comprised 6–8% of all cases and generally showed very low levels of expression spanning the threshold for positivity. These may represent a so called “borderline category” where expression cannot readily be designated into a clear-cut positive or negative status. Ideally, information on the tumour response to immunotherapy should determine how those cases should be classified. It is of interest that expert pathologists, who routinely reported PD-L1 in breast cancer, showed

substantial concordance in scoring those difficult cases. We therefore recommend that those cases of very low expression (i.e., close to the 1% cut-off value) are scored by an expert pathologist either via double-reporting or via a second opinion referral.

Table 9. Summary of studies evaluating the SP142 PD-L1 concordance of scoring.

Reference	Number of Cases (Type)	Clone(s)	SP142 Scoring Method	Scorers	Inter-Observer Agreement	Intra-Observer Agreement
Downes et al. 2020 [19]	30 surgical excisions TMAs	22C3, SP142, E1L3N	$IC \geq 1\%$	3 pathologists	Kappa for IC1%: 0.668	1 month washout period. Kappa = 0.798
Noske et al. [13]	30 (resections)	SP263, SP142, 22C3, 28–8	$IC \geq 1\%$	7 trained + one Ventana SP142 expert for SP142 only	ICC for SP142: 0.805 (0.710–0.887)	Not tested
Dennis et al. (abstract) [14]	28 test sets through the Roche International Training Programme	SP142	$IC \geq 1\%$	432 (trained multiple institutions), from several countries	OPA: was 98.2%, with PPA of 99.4% and NPA of 96.6%.	Not tested
Hoda et al. [23]	75 (cores and excision), primary and metastases	SP142	$IC \geq 1\%$	8 experienced (single institution)	Kappa 0.727	Not tested
Reisenbichler et al. 2021 [21]	68 cases for SP142 and 67 cases for SP263	SP142, SP263	$IC \geq 1\%$ & % expression for cases scored as positive only	19 randomly selected pathologists from 14 US institutions; breast pathologists, with few non-breast pathologists. Experience in reporting PD-L1 not stated	Complete agreement for SP142 categorisation into positive vs. negative in 38%. Agreement decreased with the increasing number of scorers, reaching a low plateau of 0.41 at eight scorers or more	Not tested
Pang et al. [26]	60 TNBC TMAs	VENTANA SP142, DAKO 22C3	$IC \geq 1\%$	10 pathologists including 5 PD-L1 who were naïve and 5 who passed a proficiency test	93.3% for experts; 81.5% for non-experts.	Tested after a 1 h training video and an overnight washout period. OPA increased from 81.5% to 85.7% for non-experts after video training. OPA was 96.3% for experts.
Van Bockstal et al. 2021 [15]	49 metastatic TNBC (biopsies and resections)	VENTANA SP142	$IC \geq 1\%$	10 pathologists; all passed a proficiency test	Substantial variability at the individual patient level. In 20% of cases, chance of allocation to treatment was random, with a 50–50 split among pathologists in designating as PD-L1-positive or -negative	Not tested

Table 9. Cont.

Reference	Number of Cases (Type)	Clone(s)	SP142 Scoring Method	Scorers	Inter-Observer Agreement	Intra-Observer Agreement
Ahn et al. 2021 [27]	30 surgical excisions	SP142, SP263, 22C3 and E1L3N	ICs and TCs were scored in both continuous scores (0–100%) and five categorical scores (<1%, 1–4%, 5–9%, 10–49% and ≥50%).	10 pathologists with no special training, of whom 6 underwent Ventana Roche training	80.7% inter-observer agreement at a 1% cut-off value	Proportion of cases with identical scoring at a 1% IC cut-off value increased from 40% to 70.0% after training
Abreu et al. 2022 (Conference abstract) [28]	168 in tissue microarrays	22C3 and SP142	Not stated	4 pathologists including 2 breast pathologists and 2 surgical pathologists with no specific PD-L1 training	Overall concordance for SP142 was 64.8%; overall $\kappa = 0.331$, with $\kappa = 0.420$ for breast pathologists and $\kappa = 0.285$ for general pathologists	Not tested
Chen et al. 2022 [22]	426 primary and metastatic surgical excisions	SP142	IC ≥ 1%	Two experienced pathologists	78.2% concordance; $\kappa = 0.567$	Not tested
Current study	100 (cores), primary breast cancer	SP142	IC ≥ 1% & % expression for all cases; two rounds of scoring separated by a 3-month washout period	12 experienced breast pathologists from 8 institutions in the UK, Ireland and Belgium. All passed a proficiency test.	Absolute agreement was substantial in 52% and 60% of cases in the first and second rounds, with Kappa values of 0.654 and 0.655 for the first and second rounds, respectively. Higher concordance among experts, particularly in TNBC and challenging cases.	Tested after 3 months of a washout period. Almost perfect agreement regardless of pathologists' PD-L1 experience

Similar challenges in PD-1 scoring have been highlighted in carcinomas in other tissues. For example, the concordance between the assays used for PD-L1 assessment in head and neck squamous cell carcinoma (HNSCC) was fair to moderate, with a tendency for the SP142 assay to better stain the immune cells [29]. Furthermore, using 3 PD-L1 tests for HNSCC tissue microarrays (standard SP263, standard 22C3 and in-house-developed 22C3), significant differences were found among the three tests using clinically relevant cut-off values, i.e., ≥20 and ≥50%, for the combined positive score (CPS) and Tumour positive score (TPS). Intra-tumour heterogeneity was generally higher when CPS was used [30]. On the other hand, Cerbelli et al. showed a high concordance between the 22C3 PharmDx assay and the SP263 assay on 43 whole sections of HNSCC [31]. The data collectively highlight the challenges in PD-L1 assessment in various cancers, including the differences in the results between the available antibody clones and staining platforms.

Our data also confirm previous studies showing the highest proportion of PD-L1 positivity in TNBC [32]. PD-L1 was previously shown to be associated with higher tumour grades and higher pCR rates. Low levels of expression were associated with shorter recurrence-free survival (RFS), including following subtype adjustment [32].

The current study and previous lessons from the IMpassion trial [33] shed some light on issues related to the immunohistochemical assessment of PD-L1 in breast cancer tissue. The strengths of the study include the large cohort of cases, the inclusion of 12 pathologists

from three countries, the inclusion of both expert and non-expert assessors, the robust design, with the assessment of inter- and intra-observer concordance in two rounds, and the detailed statistical analysis. The digital analysis of whole slide images, rather than scoring glass slides, may be a weakness for pathologists who are not used to digital reporting. More recently, the use of digital image analysis algorithms and/or artificial intelligence (AI) has been proposed for PD-L1 scoring in various solid tumours [34]. Going forward, this is an exciting and promising endeavour that requires thorough validation in comparison to the gold-standard pathologist scoring before implementation and the determination of whether those algorithms are superior to manual scoring in identifying responders to immune therapy. Currently, PD-L1 Artificial Intelligence (AI) scoring in breast cancer is limited to research studies and has not been validated for routine clinical use.

5. Conclusions

In summary, we present a detailed analysis of 12 pathologists who scored 100 digitally scanned breast cancer slides for PD-L using the Ventana SP142 assay in two rounds separated by a washout period. Absolute (100%) agreement was substantial in 52% and 60% of cases in the first and second rounds, with Kappa values of 0.654 and 0.655 for rounds one and two, respectively. We provide reassuring evidence of a high concordance of PD-L1 reporting among pathologists, the highest being among experts and in reporting challenging, low-expressing TNBC. The intra-observer agreement was substantial for all raters. Despite experience and the adherence to current reporting guidelines, there remains a minority of tumours (6–8%) that are challenging to assign to either a positive or negative category. Those are PD-L1 low-expressing and/or heterogeneous tumours that suffer from the least concordance among pathologists. Consensus scoring and referrals for expert opinions should be considered in those cases. If uncertainty persists, this should be recognised and well communicated to clinicians in the context of a multidisciplinary approach. For inconclusive cases, testing on another tumour sample and/or using another assay (e.g., the DAKO 22C3 assay for selecting patients for pembrolizumab therapy) could be performed.

Pathologists' training and experience are paramount in evaluating PD-L1 expression and selecting patients for immune checkpoint anti-PD-L1 inhibitors. Further work on refining the criteria for scoring, pathologists' training and assessing pathologist concordance is needed. This will ensure the accurate classification of tumours into a positive or negative category and, hence, the accurate selection of patients for atezolizumab therapy.

This study also shows that digital pathology is a useful tool that allows for the instantaneous sharing of high-quality whole slide scans with colleagues. This is particularly helpful for consensus scoring and/or seeking expert opinions.

Author Contributions: M.Z.: Virtual slide scoring, data curation, statistical analysis, results interpretation, designing tables and figures, writing; M.V.B., C.G., G.C., E.P., R.H., C.D., N.M.B., J.S., B.T. and Y.M.: Virtual slide scoring, editing the manuscript; B.O.: Slide staining; P.L.: Statistical analysis, statistical writeup, results interpretation; A.M.S.: Conceptualisation, methodology, virtual slide scoring, statistical analysis, writing, overall project overview. All authors have read and agreed to the published version of the manuscript.

Funding: M.Z. was funded by a grant from the Egyptian Cultural and Educational Bureau and sponsored by the Ministry of Higher Education and Scientific Research, General Mission Sector (Egypt). A.M.S. is supported by a Birmingham CR-UK Centre Grant, CA17422/A25154.

Institutional Review Board Statement: This study did not require ethical approval, as it was conducted as an audit of the consistency of scoring of anonymous breast cancer images.

Informed Consent Statement: Not applicable.

Data Availability Statement: Full data are available from the corresponding author upon reasonable request. Digital slides are password-protected and available at the University of Birmingham platform: <https://eslidepath.bham.ac.uk>, Last accessed 23 February 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shaaban, A.M.; Shaw, E.C. Bench to bedside: Research influencing clinical practice in breast cancer. *Diagn. Histopathol.* **2022**, *28*, 473–479. [\[CrossRef\]](#)
2. Vennapusa, B.; Baker, B.; Kowanetz, M.; Boone, J.; Menzl, I.; Bruey, J.-M.; Fine, G.; Mariathasan, S.; McCaffery, I.; Mocci, S.; et al. Development of a PD-L1 Complementary Diagnostic Immunohistochemistry Assay (SP142) for Atezolizumab. *Appl. Immunohistochem. Mol. Morphol.* **2019**, *27*, 92–100. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Melaiu, O.; Lucarini, V.; Cifaldi, L.; Fruci, D. Influence of the Tumor Microenvironment on NK Cell Function in Solid Tumors. *Front. Immunol.* **2020**, *10*, 3038. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Simiczjew, A.; Dratkiewicz, E.; Mazurkiewicz, J.; Ziętek, M.; Matkowski, R.; Nowak, D. The Influence of Tumor Microenvironment on Immune Escape of Melanoma. *Int. J. Mol. Sci.* **2020**, *21*, 8359. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Badr, N.M.; McMurray, J.L.; Daniai, I.; Hayward, S.; Asaad, N.Y.; El-Wahed, M.M.A.; Abdou, A.G.; El-Dien, M.M.S.; Sharma, N.; Horimoto, Y.; et al. Characterization of the Immune Microenvironment in Inflammatory Breast Cancer Using Multiplex Immunofluorescence. *Pathobiology* **2022**, *90*, 31–43. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Conte, P.F.; Dieci, M.V.; Bisagni, G.; De Laurentiis, M.; Tondini, C.A.; Schmid, P.; De Salvo, G.L.; Moratello, G.; Guarneri, V. Phase III randomized study of adjuvant treatment with the ANTI-PD-L1 antibody avelumab for high-risk triple negative breast cancer patients: The A-BRAVE trial. *Am. Soc. Clin. Oncol.* **2020**, *38*, TPS598. [\[CrossRef\]](#)
7. Schmid, P.; Cortes, J.; Dent, R. Pembrolizumab in Early Triple-Negative Breast Cancer. Reply. *N. Engl. J. Med.* **2022**, *386*, 1771–1772. [\[CrossRef\]](#)
8. Emens, L.A.; Adams, S.; Barrios, C.H.; Diéras, V.; Iwata, H.; Loi, S.; Rugo, H.S.; Schneeweiss, A.; Winer, E.P.; Patel, S.; et al. First-line atezolizumab plus nab-paclitaxel for unresectable, locally advanced, or metastatic triple-negative breast cancer: IMpassion130 final overall survival analysis. *Ann. Oncol.* **2021**, *32*, 983–993. [\[CrossRef\]](#)
9. Bartsch, R. ESMO 2020: Highlights in breast cancer. *Memo-Mag. Eur. Med. Oncol.* **2021**, *14*, 184–187. [\[CrossRef\]](#)
10. Büttner, R.; Gosney, J.R.; Skov, B.G.; Adam, J.; Motoi, N.; Bloom, K.J.; Dietel, M.; Longshore, J.W.; López-Ríos, F.; Penault-Llorca, F.; et al. Programmed Death-Ligand 1 Immunohistochemistry Testing: A Review of Analytical Assays and Clinical Implementation in Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2017**, *35*, 3867–3876. [\[CrossRef\]](#)
11. Guo, H.; Ding, Q.; Gong, Y.; Gilcrease, M.Z.; Zhao, M.; Zhao, J.; Sui, D.; Wu, Y.; Chen, H.; Liu, H.; et al. Comparison of three scoring methods using the FDA-approved 22C3 immunohistochemistry assay to evaluate PD-L1 expression in breast cancer and their association with clinicopathologic factors. *Breast Cancer Res.* **2020**, *22*, 69. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Huang, X.; Ding, Q.; Guo, H.; Gong, Y.; Zhao, J.; Zhao, M.; Sui, D.; Wu, Y.; Chen, H.; Liu, H.; et al. Comparison of three FDA-approved diagnostic immunohistochemistry assays of PD-L1 in triple-negative breast carcinoma. *Hum. Pathol.* **2021**, *108*, 42–50. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Noske, A.; Wagner, D.-C.; Schwamborn, K.; Foersch, S.; Steiger, K.; Kiechle, M.; Oettler, D.; Karapetyan, S.; Hapfelmeier, A.; Roth, W.; et al. Interassay and interobserver comparability study of four programmed death-ligand 1 (PD-L1) immunohistochemistry assays in triple-negative breast cancer. *Breast* **2021**, *60*, 238–244. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Dennis, E.; Kockx, M.; Harlow, G.; Cai, Z.; Bloom, K.; ElGabry, E. Abstract PD5-02: Effective and globally reproducible digital pathologist training program on PD-L1 immunohistochemistry scoring on immune cells as a predictive biomarker for cancer immunotherapy in triple negative breast cancer. *Cancer Res.* **2020**, *80* (Suppl. 4), PD5-02. [\[CrossRef\]](#)
15. Van Bockstal, M.R.; Cooks, M.; Nederlof, I.; Brinkhuis, M.; Dutman, A.; Koopmans, M.; Kooreman, L.; van der Vegt, B.; Verhoog, L.; Vreuls, C.; et al. Interobserver Agreement of PD-L1/SP142 Immunohistochemistry and Tumor-Infiltrating Lymphocytes (TILs) in Distant Metastases of Triple-Negative Breast Cancer: A Proof-of-Concept Study. A Report on Behalf of the International Immuno-Oncology Biomarker Working Group. *Cancers* **2021**, *13*, 4910.
16. Padmanabhan, R.; Kheraldine, H.S.; Meskin, N.; Vranic, S.; Al Moustafa, A.-E. Crosstalk between HER2 and PD-1/PD-L1 in breast cancer: From clinical applications to mathematical models. *Cancers* **2020**, *12*, 636. [\[CrossRef\]](#)
17. Kurozumi, S.; Inoue, K.; Matsumoto, H.; Fujii, T.; Horiguchi, J.; Oyama, T.; Kurosumi, M.; Shirabe, K. Clinicopathological values of PD-L1 expression in HER2-positive breast cancer. *Sci. Rep.* **2019**, *9*, 16662. [\[CrossRef\]](#)
18. Marletta, S.; Fusco, N.; Munari, E.; Luchini, C.; Cimadamore, A.; Brunelli, M.; Querzoli, G.; Martini, M.; Vigliar, E.; Colombari, R.; et al. Atlas of PD-L1 for Pathologists: Indications, Scores, Diagnostic Platforms and Reporting Systems. *J. Pers. Med.* **2022**, *12*, 1073. [\[CrossRef\]](#)
19. Downes, M.R.; Slodkowska, E.; Katabi, N.; Jungbluth, A.; 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*, 191–200. [\[CrossRef\]](#)
20. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [\[CrossRef\]](#)
21. Reisenbichler, E.S.; Han, G.; Bellizzi, A.; Bossuyt, V.; Brock, J.; Cole, K.; Fadare, O.; Hameed, O.; Hanley, K.; Harrison, B.T.; 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*, 1746–1752. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Chen, C.; Ma, X.; Li, Y.; Ma, J.; Yang, W.; Shui, R. Concordance of PD-L1 expression in triple-negative breast cancers in Chinese patients: A retrospective and pathologist-based study. *Pathol. Res. Pract.* **2022**, *238*, 154137. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Hoda, R.S.; Brogi, E.; D'Alfonso, T.M.; Grabenstetter, A.; Giri, D.; Hanna, M.G.; Kuba, M.G.; Murray, D.M.P.; Vallejo, C.E.; Zhang, H.; et al. Interobserver Variation of PD-L1 SP142 Immunohistochemistry Interpretation in Breast Carcinoma: A Study of 79 Cases Using Whole Slide Imaging. *Arch. Pathol. Lab. Med.* **2021**, *145*, 1132–1137. [[CrossRef](#)] [[PubMed](#)]
24. A Rakha, E.; E Pinder, S.; Bartlett, J.M.S.; Ibrahim, M.; Starczynski, J.; Carder, P.J.; Provenzano, E.; Hanby, A.; Hales, S.; Lee, A.H.S.; et al. Updated UK Recommendations for HER2 assessment in breast cancer. *J. Clin. Pathol.* **2014**, *68*, 93–99. [[CrossRef](#)]
25. Peg, V.; López-García, M.; Comerma, L.; Peiró, G.; García-Caballero, T.; López, Á.C.; Suárez-Gauthier, A.; Ruiz, I.; Rojo, F. PD-L1 testing based on the SP142 antibody in metastatic triple-negative breast cancer: Summary of an expert round-table discussion. *Future Oncol.* **2021**, *17*, 1209–1218. [[CrossRef](#)]
26. Pang, J.-M.B.; Castles, B.; Byrne, D.J.; Button, P.; Hendry, S.; Lakhani, S.R.; Sivasubramaniam, V.; Cooper, W.A.; Armes, J.; Millar, E.K.; et al. SP142 PD-L1 Scoring Shows High Interobserver and Intraobserver Agreement in Triple-negative Breast Carcinoma But Overall Low Percentage Agreement With Other PD-L1 Clones SP263 and 22C3. *Am. J. Surg. Pathol.* **2021**, *45*, 1108–1117. [[CrossRef](#)]
27. Ahn, S.; Woo, J.W.; Kim, H.; Cho, E.Y.; Kim, A.; Kim, J.Y.; Kim, C.; Lee, H.J.; Lee, J.S.; Bae, Y.K.; et al. Programmed Death Ligand 1 Immunohistochemistry in Triple-Negative Breast Cancer: Evaluation of Inter-Pathologist Concordance and Inter-Assay Variability. *J. Breast Cancer* **2021**, *24*, 266–279. [[CrossRef](#)]
28. Abreu, R.; Peixoto, Corassa, M.; Nunes, W.; Neotti, T.; Rodrigues, T.; Toledo, C.; Domingos, T.; Carraro, D.; Gobbi, H.; et al. Determination of Inter-Observer Agreement in the Immunohistochemical Interpretation of PD-L1 Clones 22C3 and SP142 in Triple-Negative Breast Cancer (TNBC). In *Virchows Archiv*; Springer One New York Plaza; Springer: New York, NY, USA, 2022.
29. Girolami, I.; Pantanowitz, L.; Barberis, M.; Paolino, G.; Brunelli, M.; Vigliar, E.; Munari, E.; Satturwar, S.; Troncone, G.; Eccher, A. Challenges facing pathologists evaluating PD-L1 in head & neck squamous cell carcinoma. *J. Oral Pathol. Med.* **2021**, *50*, 864–873.
30. de Ruiter, E.J.; Mulder, F.; Koomen, B.; Speel, E.-J.; van den Hout, M.; de Roest, R.; Bloemena, E.; Devriese, L.; Willems, S. Comparison of three PD-L1 immunohistochemical assays in head and neck squamous cell carcinoma (HNSCC). *Mod. Pathol.* **2021**, *34*, 1125–1132. [[CrossRef](#)]
31. Cerbelli, B.; Girolami, I.; Eccher, A.; Costarelli, L.; Taccogna, S.; Scialpi, R.; Benevolo, M.; Lucante, T.; Alò, P.L.; Stella, F.; et al. Evaluating programmed death-ligand 1 (PD-L1) in head and neck squamous cell carcinoma: Concordance between the 22C3 PharmDx assay and the SP263 assay on whole sections from a multicentre study. *Histopathology* **2021**, *80*, 397–406. [[CrossRef](#)]
32. Vidula, N.; Yau, C.; Rugo, H.S. Programmed cell death 1 (PD-1) receptor and programmed death ligand 1 (PD-L1) gene expression in primary breast cancer. *Breast Cancer Res. Treat.* **2021**, *187*, 387–395. [[CrossRef](#)] [[PubMed](#)]
33. Badve, S.S.; Penault-Llorca, F.; Reis-Filho, J.S.; Deurloo, R.; Siziopikou, K.P.; D'Arrigo, C.; Viale, G. Determining PD-L1 Status in Patients With Triple-Negative Breast Cancer: Lessons Learned From IMpassion130. *Gynecol. Oncol.* **2021**, *114*, 664–675. [[CrossRef](#)] [[PubMed](#)]
34. Inge, L.; Dennis, E. Development and applications of computer image analysis algorithms for scoring of PD-L1 immunohistochemistry. *Immunooncol. Technol.* **2020**, *6*, 2–8. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.