Virchows Archiv (2019) 475:67–76
https://doi.org/10.1007/s00428-019-02582-0

ORIGINAL ARTICLE

PD-L1 testing of non-small cell lung cancer using different antibodies and platforms: a Swiss cross-validation study

Spasenija Savic[1] · Sabina Berezowska[2] · Serenella Eppenberger-Castori[1] · Gieri Cathomas[3] · Joachim Diebold[4] · Achim Fleischmann[5] · Wolfram Jochum[6] · Paul Komminoth[7] · Thomas McKee[8] · Igor Letovanec[9] · Zerina Jasarevic[10] · Matthias Rössle[11,12] · Gad Singer[13] · Michael von Gunten[14] · Andreas Zettl[15] · Roland Zweifel[16] · Alex Soltermann[17] · Lukas Bubendorf[1]

Received: 1 November 2018 / Revised: 12 April 2019 / Accepted: 7 May 2019 / Published online: 24 May 2019

Abstract

With the approval of pembrolizumab for first- and second-line treatment of PD-L1+ non-small cell lung cancer (NSCLC), PD-L1 testing by immunohistochemistry (IHC) has become a necessity. However, the DAKO autostainer ASL48 for the FDA approved DAKO 22C3 pharmDx assay is not broadly available in Switzerland and other parts of Europe. The primary goal of this study was to cross-validate the 22C3 anti-PD-L1 antibody on Benchmark Ultra (VBMU) and Leica Bond (LBO) immunostainers. IHC protocols were developed for 22C3 on both platforms with the 22C3phDx using ASL48 as reference. A tissue microarray (TMA) was constructed from 23 NSCLC specimens with a range of PD-L1 staining results. Empty TMA sections and the 22C3 antibody were distributed to 16 participants for staining on VBMU (8 centers) and/or LBO (12 centers) using the centrally developed protocols. Additionally the performance of the Ventana SP263 assay was tested in five centers. IHC scoring was performed centrally. Categorical PD-L1 staining (0–49% vs. 50–100%) did not significantly differ between centers using VBMU, whereas data from LBO were highly variable ($p < 0.001$). The SP263 assay was well concordant with 22C3 on VBMU and with 22C3 pharmDx. PD-L1 IHC using a standardized 22C3 protocol on VBMU provides satisfactory results in most centers. The SP263 assay is confirmed as a valid alternative to 22C3 pharmDx. 22C3 PD-L1 IHC on LBO shows major staining variability between centers, highlighting the need for local validation and adjustment of protocols.

Keywords Non-small cell lung carcinoma · PD-L1 · Immunohistochemistry · 22C3 · SP263

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00428-019-02582-0) contains supplementary material, which is available to authorized users.

✉ Lukas Bubendorf
Lukas.Bubendorf@usb.ch

[1] Institute of Medical Genetics and Pathology, University Hospital Basel, Schönbeinstrasse 40, 4031 Basel, Switzerland

[2] Institute of Pathology, University of Bern, Bern, Switzerland

[3] Cantonal Hospital Baselland, Institute of Pathology, Liestal, Switzerland

[4] Cantonal Hospital Lucerne, Institute of Pathology, Lucerne, Switzerland

[5] Institute of Pathology, Kantonsspital Münsterlingen, Münsterlingen, Switzerland

[6] Institute of Pathology, Kantonsspital St. Gallen, St. Gallen, Switzerland

[7] Department of Pathology, Triemli Hospital Zurich, Zurich, Switzerland

[8] Institute of Pathology, University Geneva, 1211 Geneva, Switzerland

[9] Service of Pathology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

[10] Department of Pathology, Landeskrankenhaus Feldkirch, Feldkirch, Austria

[11] Cantonal Hospital Graubünden, Institute of Pathology, Chur, Switzerland

[12] Present address: University of Zurich, Zurich, Switzerland

[13] Institute of Pathology, Kantonsspital Baden AG, Baden, Switzerland

[14] Länggasse Bern, Institute of Pathology, Bern, Switzerland

[15] Viollier AG, Allschwil, Switzerland

[16] Institute of Pathology, Cantonal Hospital Aarau, Aarau, Switzerland

[17] Institute of Pathology, University Hospital Zurich, Zurich, Switzerland

Introduction

Immunotherapy by checkpoint blockade targeting programmed death 1 (PD-1) or its ligand programmed death ligand 1 (PD-L1) has become a standard of treatment in non-small cell lung cancer (NSCLC). Five PD-1 (pembrolizumab and nivolumab) or PD-L1 inhibitors (atezolizumab, durvalumab, and avelumab) have been approved for first- and/or second-line treatment. The response of NSCLC to these agents is often associated with the extent of immunohistochemical PD-L1 expression in tumor cells and/or associated immune cells. Immunohistochemistry (IHC) for diagnostic and predictive marker testing is a well-established method in general. However, lack of standardized PD-L1 testing because of different clinical-trial validated PD-L1 assays, and different, indication-specific scoring algorithms have led to considerable confusion. Other challenges include tumor heterogeneity, variability in pre-analytical conditions, and inter-observer variability. The different commercially available PD-L1 IHC assays have been compared against each other in a number of studies [1–5]. This has provided important insights as to how the different assays linked to specific immunostainer platforms compare against each other. Notably, it has been shown that the Ventana SP263 assay (Ventana, Tucson, AZ) on the widely used Ventana Benchmark immunostainer (Ventana, Tucson, AZ) is highly concordant to the DAKO 22C3 and 28-8 pharmDx assays (Agilent, Santa Clara, CA), which were optimized for the DAKO Autostainer Link 48 (ASL48; Dako, Carpinteria, CA) [2, 3, 5]. For this reason, the SP263 assay has recently been Communauté Européenne (CE)-marked for interchangeable use with the pembrolizumab-related 22C3 and the nivolumab-related 28-8 pharmDx assays [6]. In addition to these relatively expensive PD-L1 assays, many institutions use laboratory developed tests (LDT) with a variety of antibodies, including 22C3 and the less expensive E1L3N antibody (Cell Signaling Technology, Danvers, MA) for the purpose of saving costs or because the respective immunostainer platform linked to a respective assay is not available in the laboratory. Notably, there are no clinically validated commercial PD-L1 assays for the Leica Bond platform (LBO) (Leica Biosystems, Wetzlar, Germany), which is also widely used.

Since the approval of pembrolizumab monotherapy for first-line treatment of advanced, PD-L1-positive NSCLC, pathologists have to deliver systematic PD-L1 IHC testing. In the first-line setting, pembrolizumab is most effective in patients with a high proportion of ≥ 50% of PD-L1-positive tumor cells (TCs) [7, 8]. This PD-L1 expression threshold of 50% is therefore a prerequisite for treatment. In the pivotal trials, the DAKO PD-L1 antibody 22C3 has been used on the DAKO ASL48 for predictive IHC. Based on these data, the Federal Drug Administration (FDA) of the USA approved the drug in patients with PD-L1-positive tumors determined by this particular IHC assay (DAKO 22C3 pharmDx). However, the DAKO platform is not widely available in Switzerland (only in one institute) and other European countries, and the Ventana Benchmark or the LBO platforms are used more commonly. This creates an issue in systematic predictive PD-L1 testing in order to select patients for pembrolizumab if usage of the FDA-approved DAKO assay linked to the DAKO platform is considered mandatory. The European Medicine Agency (EMA) and the Swiss agency Swissmedic are less stringent than the FDA, as they do not prescribe a specific PD-L1 IHC assay. Nevertheless, there is an apparent need to explore if the DAKO PD-L1 antibody 22C3, which has been used in the pivotal pembrolizumab clinical trials, is also applicable to other automated immunostainer platforms using the DAKO assay on the DAKO ASL48 as a gold standard for comparison. In contrast to ready-to-use IHC assay kits, establishing a laboratory developed test (LDT) requires rigorous protocol development and validation with appropriate positive controls, which can be challenging for individual laboratories. Here, we centrally developed and validated new PD-L1 IHC protocols on the Ventana Benchmark Ultra (VBMU) and LBO using concentrated 22C3 antibody and tested their concordance across different laboratories in Switzerland. In addition, we explored the interlaboratory concordance of the SP263 assay in centers, where this assay was already available.

Materials and methods

Immunohistochemistry protocol development

IHC protocol development and PD-L1 staining using 22C3 pharmDX as the reference standard were performed centrally at the Institute of Pathology (IfP), University Hospital Basel. For the 22C3 pharmDX assay, the ASL48 automated immunostainer was rented from DAKO for the purpose of this study. IHC protocols for the LDT were established using concentrated 22C3 antibody (Dako) for the VBMU and for the LBO platforms, respectively. PD-L1-positive control material for protocol establishment included formalin-fixed and paraffin-embedded (FFPE) specimens from normal tonsil and placenta after overnight fixation in 10% buffered formalin, and cellblocks from the Karpas cell line (diffuse, moderately strong PD-L1 staining) and the prostate cancer cell line LNCap (focal, weak PD-L1 staining). Cellblocks were prepared using the plasma-thrombin method [9]. FFPE tissue sections from selected resection specimens of NSCLC with known PD-L1 IHC status showing a range of PD-L1 expression results were used as additional positive and negative controls. Only PD-L1 tumor proportion score (TPS) was considered in this study, since immune cell score (ICS) is irrelevant for first-line treatment with pembrolizumab [8]. TPS represents the estimated percentage (0–100%) of TCs showing

partial or complete membranous PD-L1 staining [10]. Two pathologists (S.S. and L.B.) evaluated the staining and selected the protocol for VBMU and LBO, respectively, that were closest to the results of the 22C3 pharmDX reference (detailed protocols in the supplementary material). Overall, IHC staining characteristics between the reference and the selected protocol for the concentrated 22C3 on VBMU were considered to be practically identical. In contrast, the IHC staining of the concentrated 22C3 antibody on the LBO platform was generally weaker and showed a lighter brown as compared to the other two platforms, often requiring a higher level of magnification for assigning the IHC score. The selected IHC protocols for the 22C3-based LDTs on VBMU and LBO were then independently applied on the same set of test material at the Institute of Pathology, University Hospital Zurich, and technical reproducibility of the PD-L1 staining results was confirmed at a joint meeting (S.S., LB., and A.S.).

Tissue microarray construction

Representative FFPE blocks from surgically resected NSCLC with known PD-L1 IHC status were selected by the University IfP in Basel ($n = 16$), Bern ($n = 4$), and Zurich ($n = 3$). A TMA with NSCLC covering a range of PD-L1 TPS and expression intensities was constructed at the IfP in Basel using the automated tissue micro arrayer TMA Grand Master (3DHISTECH Ltd., Budapest, Hungary). The TMA contains 23 NSCLC each represented by three TMA cores at 1-mm diameter and control specimens consisting of Karpas cell line, tonsil, and placenta tissue. A hematoxylin and eosin stained slide of the study TMA was scanned and made available to participants online as a virtual morphological reference slide (see URL for access in online supplementary material).

Multi-institutional cross-validation of PD-L1 IHC

All together, 16 institutions participated in this multi-institutional cross-comparison study, including 15 institutes from Switzerland and one institute from Austria. The central laboratory in Basel represents "participant 1" in this study, whereas the other participant have been anonymized. All participants were provided with the two LDT protocols together with a vial of concentrated 22C3 antibody from the same production batch, empty sections of FFPE tonsil tissue, and commercial FFPE PD-L1 control slides containing four different cell lines with a dynamic range of PD-L1 expression levels (negative, low, medium, and high) (PD-L1 Analyte Control^{DR} HistoCyte Laboratories Ltd., Newcastle, UK) for local protocol testing. Participating pathologists were also asked to use own local tumor material to assess the technical viability of the protocols under local conditions prior to TMA staining. Each participant received two freshly cut empty TMA sections per type of automated immunostainer. The automated immunostainer platforms used by the participants were LBO (12×), VBMU (8×), and both LBO and VBMU (4×). All of these automated immunostainers were in routine use for IHC service in the respective institutes. The reference TMA section was stained by the Dako 22C3 pharmDx assay on the Dako ASL48 at the IfP Basel. Additionally, five participants stained an extra TMA section with their locally established Ventana SP263 assay, which was performed according to the manufacturer's recommendations. All immunostained TMA sections were sent to the IfP Basel and independently scored by three experienced pathologists (S.S., L.B., and S.B.) who had previously received DAKO PD-L1 IHC interpretation training independent of the study. Each tissue spot was considered as a single specimen for statistical analysis, resulting in up to 76 specimens and data points. Variation in number of spots was due to mechanical loss or absence of sufficient tumor cells, and at least 12 of the 76 cores on the TMA sections were not informative. At least 100 tumor cells were required for a spot to be informative. TPS was determined as a continuous variable for each TMA spot, and the mean TPS of the three central reviewers was used for statistical analysis.

Statistical analysis

Consistency of the data obtained by the three central expert pathologists was checked for the percentage of the detected PD-L1-positive TC. Analyses were primarily based on continuous PD-L1 percentages. Mean values of the three continuous data sets were calculated all over and then transformed categorically into the three groups of interest for treatment efficacy prediction: TPS < 1 (group 1), TPS 1–49 (group 2), and TPS ≥ 50 (group 3).

Fleiss Kappa values for concordance were calculated for continuous values and data after classification into the three categories (TPS < 1; 1–49; 50–100) using the package irr and the kappam.fleiss function. Mean values of the three data sets received were calculated all over. Analysis of variance was performed to compare the mean of PD-L1-positive percentage of cells obtained under the several conditions. Tukey plots were used to visualize the ANOVA results and notched box plots to visualize the differences in median values. Significant differences in the variance of the detected values were also calculated by means of the Kruskal-Wallis test. Moreover, distribution plots were created to check the single spot variation. All p values were two-sided and considered significant when $p < 0.05$. Analyses were performed by means of the Statistical Package Software R (Version 3.4.1, (2017-06-30) www.r-project.org or higher).

Results

Inter-observer variability

Inter-observer analysis of PD-L1 scoring by the three central pathologists showed moderate to good agreement for 22C3 pharmDx, 22C3 on VBMU and 22C3 on LBO when TPS was used as a continuous variable (Table 1). When TPS was stratified into clinically meaningful categorical data as three groups (< 1; 1–49; ≥ 50), there was very good agreement for 22C3 on VBMU and for SP263 and good agreement for 22C3 pharmDx on ASL48 and22C3 on LBO (Table 1, $p < 0.0001$ for all kappa values). Representative images of PD-L1 IHC with different protocols and antibodies and across different participants are shown in Figs. 1 and 2, respectively.

Multi-institutional cross-comparison

Comparison of the average TPS values as detected by the three readers across all TMA specimens and participating institutions revealed a high degree of variability between centers (Fig. 3). The median value of the TPS detected in 12 centers using 22C3 on LBO was significantly lower as compared to one of the 8 centers using 22C3 on VBMU (Fig. 3, filled light blue and filled yellow box, respectively) ($p < 0.01$).

The difference of the median TPS between the reference 22C3 pharmDx and the average median TPS of all 22C3 on VBMU was not significant (Fig. 3, filled green and yellow box, respectively), though there were two outliers with very low median TPS values (centers 10 and 12). The average median TPS of SP263 was highly concordant with the reference 22C3 pharmDx (Fig. 3 filled blue box).

Analysis of TPS grouped into three clinically meaningful TPS categories (< 1, 1–49; ≥ 50) showed a clear shift towards a higher percentage of negative cases (TPS < 1) and a lower proportion of cases with TPS ≥ 50 in the centers using 22C3 on LBO than in those using 22C3 on VBMU (Table 2, Fig. 4). When using only two categories (0–49; ≥ 50), 22C3 on LBO showed significantly less cases with a treatment-relevant PD-L1 high level expression (TPS ≥ 50) compared to the 22C3 pharmDx reference ($p < 0.01$), whereas there was only a trend towards this direction in the centers using 22C3 on VBMU. On average, 6 (21%), 14 (48%), and 3 (10%) TMA tumor specimens would have been misclassified as PD-L1 negative for first-line treatment with pembrolizumab (TPS < 50) in the 22C3 on VBMU, 22C3 on LBO, and the SP263 group, respectively. Despite the high variability between the centers with 22C3 on LBO, some did reach a high concordance of more than 80% agreement with the reference 22C3 pharmDx assay on ASL48.

Discussion

This Swiss national cross-validation study of PD-L1 staining was born out of the urgent need to enable reliable PD-L1 testing to select NSCLC patients for first-line monotherapy with the PD-1 inhibitor pembrolizumab despite unavailability of the DAKO ASL48 platform needed for the FDA-approved companion test 22C3 pharmDx. To the best of our knowledge, this is the first multi-institutional study comparing the performance of centrally developed and thus standardized PD-L1 LDTs across different institutions. Our study illustrates the challenges and limitations of adopting central LDTs, notably without local adjustments, for PD-L1 IHC. The focus was on analytical cross-platform harmonization. Inter-observer variability between the three pathologists involved in scoring was not a relevant confounding factor given the good or very good concordance for the clinically relevant categories and cut-offs across all analyses. We show that LDTs using concentrated 22C3 on other platforms can principally provide results that are concordant with the reference DAKO 22C3 pharmDx assay on the DAKO ASL48.

The median values of the continuous TPS in the eight centers with the PD-L1 LDT using concentrated DAKO 22C3 antibody on VBMU tended to be lower but were not significantly different from the 22C3 pharmDx assay. Thus, in most centers, the centrally developed 22C3 protocol on VBMU showed a good concordance with 22C3 pharmDx on ASL48. However, there were two outlier centers with worryingly low median TPS. In these two centers, on average, 21% of specimens scored above the TPS of 50 by the 22C3 pharmDx assay fell below this treatment relevant threshold.

Previous single institutional studies using the prediluted 22C3 antibody provided in the Dako pharmDx PD-L1 assay on the VBMU platform showed good concordance with the 22C3 pharmDx assay [11, 12]. The study by Neuman et al. optimized multiple IHC conditions resulting in 22 different

Table 1 Inter-observer comparison of independent PD-L1 scoring by three central pathologists (Fleiss kappa values)

	Continuous values	Categorical values TPS < 1; 1–49; ≥ 50
22C3 Benchmark U (mean of 8 centers)	0.617	0.876
22C3 Bond (mean of 12 centers)	0.485	0.771
SPS263 Benchmark U (mean of 5 centers)	0.440	0.836

TPS tumor proportion score

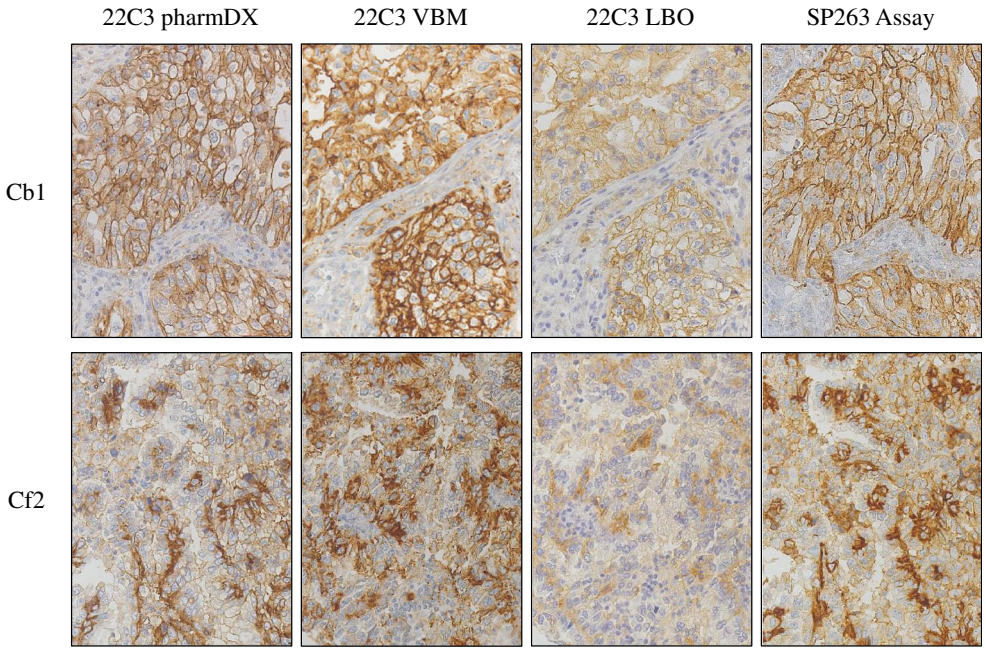

Fig. 1 PD-L1 staining results on matched areas with different protocols and in two different NSCLC specimens (all stainings performed at center 1): Core Cb1 shows PD-L1 staining in 100% of tumor cells irrespective of the protocol. Assays 22C3pharmDX and SP263 as well as 22C3 VBM show comparable staining intensities, whereas 22C3 LBO demonstrates fainter staining. Core Cf2 22C3 LBO not only stains fainter but also stains significantly fewer tumor cells (TPS < 50%). Assays 22C3pharmDX and SP263 as well as 22C3 VBM show a TPS > 50%. NSCLC non-small cell lung cancer, VBM Ventana BenchMark, LBO Leica Bond, TPS tumor proportion score

VBMU protocols, which highlights the efforts needed to establish a local LDT [11]. They selected the best protocol in relation to the 22C3 pharmDx assay for harmonization on 41 NSCLC specimens, which resulted in concordant results in 88% of cases when using categorized PD-L1 results. Similarly to our study, two studies established LDTs with concentrated 22C3 for the VBMU and found an almost 100% concordance rate for categorized TPS with the pharmDx assay when analyzed in their own laboratory, but again, they did not challenge their protocol by an inter-laboratory comparison [13, 14].

The LBO platform is in common use in Switzerland and other European countries, but has never been validated for PD-L1 testing within a clinical trial. Thus, unless these

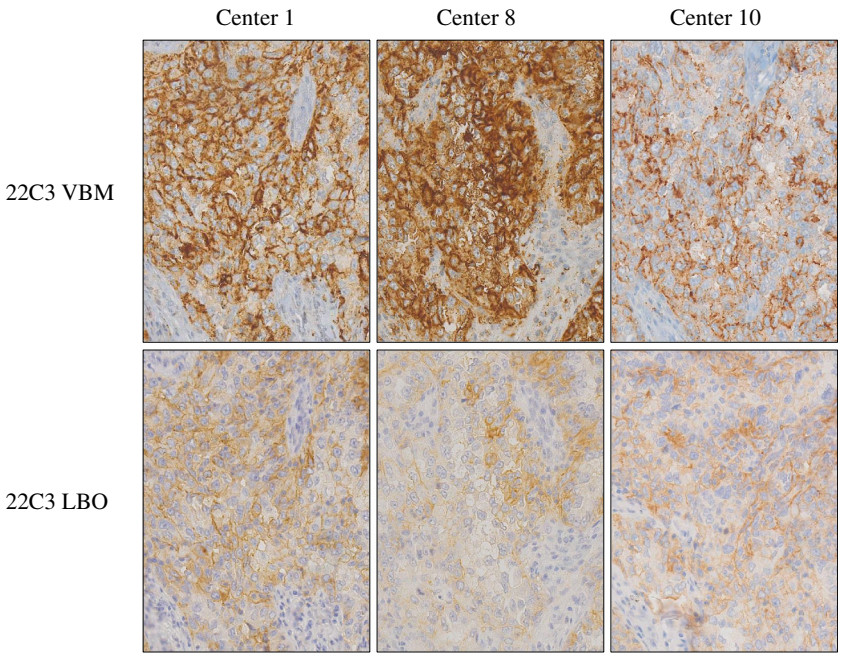

Fig. 2 Representative images of PD-L1 staining results on matched areas of core Cf1 with 22C3 VBM and 22C3 LBO, respectively, across three different centers. The images highlight the fainter staining of the LBO protocol despite some variability in the 22C3 VBM protocol. For all three centers, the 22C3 VBM shows a TPS > 50%. For 22C3 LBO, center 8 has a discordant result with a TPS < 50%. VBM Ventana Benchmark, LBO Leica Bond

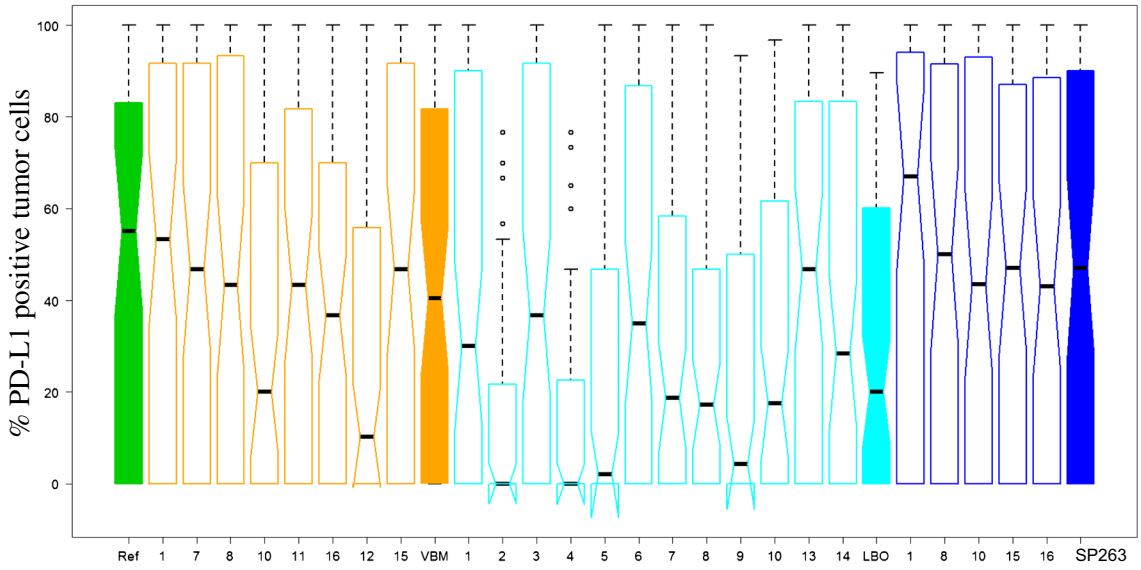

Fig. 3 Notched boxplots depicting percentage of positive tumor cells by center for the four PD-L1 protocols compared to the DAKO 22C3phDx reference standard. Reference standard (green); 22C3 on VBM in 8 centers and their mean (filled box) (yellow); 22C3 on LBO in 12 centers and their mean (filled box) (light blue); and SP263 in 5 centers and their mean (filled box) (dark blue). VBM Ventana Benchmark, LBO Leica Bond

laboratories have a second platform on which they can run a clinically validated PD-L1 assay or send out their specimens for PD-L1 testing, they have to rely on LDTs. 22C3 LDT on LBO was clearly more challenging given the significant discordance between the 12 centers and overall low mean TPS. Notably, on average, almost half of all samples with a TPS ≥ 50 by the reference 22C3pharmDx assay would have been false-negative by this cut-off. Nevertheless, it is remarkable that few laboratories reached satisfactory results, illustrating that reliable testing on LBO is possible under optimal analytical conditions.

Our findings show that centrally developed protocols can provide valid PD-L1 results, but that local optimization of protocols is crucial for valid PD-L1 testing as laboratory-specific conditions may significantly differ (for example, platform calibration, water conditions, etc.). Additionally, the 22C3 on LBO seems to be more sensitive to local conditions than on VBMU. The study by Roge et al. also demonstrated uniformly lower TPS values on the Leica platform compared to the reference 22C3pharmDx assay and a LDT using 22C3 on VBMU, despite efforts to improve the protocol. However, this had an impact on only one of 14 NSCLC (7%) with TPS ≥ 50 by the reference assay, which was misclassified as PD-L1 negative by the Leica protocol [14].

A comparative interlaboratory comparison of the 22C3pharmDx reference assay would have been interesting in order to test the robustness of this highly standardized, ready-to-use test. Unfortunately, due to the unavailability of the DAKO ASL48 platform, we were not able to perform such an analysis. Two previous multi-institutional studies tested the 22C3pharmDx between three institutions each with concordance coefficients in the substantial to almost perfect range (kappa values for three step scoring TPS1, TPS1–49, and TPS ≥ 50: 0.82–0.91) [4, 5]. The robustness of the assay was also evident in the PD-L1 module of the international external quality assurance program NordiQC, where the assay provided an overall pass rate of 96% (27 of 28 institutions) [15]. During the course of our study, the Ventana SP263 PD-L1 assay was CE-marked and cleared to be interchangeable with the DAKO 22C3 pharmDx assay, providing another valid option for PD-L1 testing for laboratories having only the Ventana platform in use [6]. This decision is supported by the high concordance between the SP263 and the 22C3 pharmDx assays in previous studies and also evident from our own data [1, 3–5, 16, 17]. The SP263 assay applied according to the manufacturer's recommendations showed a high concordance between the five institutions and was almost equivalent to DAKO 22C3 pharmDx on ASL48. Notably, it has recently been shown that the 22C3 pharmDx and SP263 assays did not only have a good analytical concordance at a TPS ≥ 50 cut-off but also an equivalent predictive performance in patients treated with nivolumab [18]. These rather uniform findings have recently been challenged by others questioning the interchangeability of 22C3 and SP263 PD-L1 assays [12, 19]. For example, Munari et al. found twice as many PD-L1-positive NSCLC specimens with SP263 on Ventana than with 22C3 both on the Ventana and ASL48 platform at the TPS ≥ 50 cut-off [19]. Conversely, the higher positivity rate of SP263 was restricted to the cut-off of TPS ≥ 1 by Hendry et al. This controversy emphasizes the persistent challenge of immunohistochemical PD-L1 testing due to a multitude of factors including biological tumor heterogeneity, variation in

Table 2 Prevalence of categorical PD-L1 results by protocol and center

PD-L1 protocol	Center	PD-L1 Tumor proportion score						Total *n*
		< 1		(1–49)		≥ 50		
		n	%	*n*	%	*n*	%	
22C3 PhDX	1	15	27.8	10	18.5	29	53.7	54
22C3 VBM	1	24	38.1	6	9.5	33	52.4	63
	7	24	38.1	8	12.7	31	49.2	63
	8	23	37.7	8	13.1	30	49.2	61
	10	26	40.6	15	23.4	23	35.9	64
	11	25	39.7	12	19.0	26	41.3	63
	12	26	41.3	11	17.5	26	41.3	63
	15	27	42.9	17	27.0	19	30.2	63
	16	21	32.8	12	18.8	31	48.4	64
	Mean (± SD)	24 (± 1.9)	37.5	12 (± 3.7)	18.8	28 (± 4.7)	43.8	64
22C3 LBO	1	25	39.7	14	22.2	24	38.1	63
	2	36	59.0	15	24.6	10	16.4	61
	3	24	37.5	11	17.2	29	45.3	64
	4	40	62.5	19	29.7	5	7.8	64
	5	29	46.0	19	30.2	15	23.8	63
	6	22	34.9	13	20.6	28	44.4	63
	7	28	44.4	16	25.4	19	30.2	63
	8	27	44.3	19	31.1	15	24.6	61
	9	29	45.3	18	28.1	17	26.6	64
	10	26	40.6	17	26.6	21	32.8	64
	13	18	31.0	11	19.0	29	50.0	58
	14	24	37.5	13	20.3	27	42.2	64
	Mean (± SD)	27 (± 5.9)	42.2	18 (± 3.0)	28.1	19 (± 7.8)	29.7	64
SP263 assay	1	16	29.1	9	16.4	30	54.5	55
	8	16	31.4	9	17.6	26	51.0	51
	10	17	31.5	11	20.4	26	48.1	54
	15	17	30.9	11	20.0	27	49.1	55
	16	17	30.9	12	21.8	26	47.3	55
	Mean (± SD)	16 (± 0.5)	29.6	10 (± 1.3)	18.5	28 (± 1.7)	51.9	54

analytical chemistry, antibody clones, financial factors, and unavoidable inter-observer variability [12]. It has recently been emphasized that PD-L1 IHC remains an imperfect biomarker and that it is unfair to compare it with addictive oncogene biomarkers that can stratify patients for targeted treatment more precisely and accurately by nature [18, 19].

This study has certain limitations. The TMA contained a relatively low number of 23 NSCLC specimens. However, by using up to three samples per tumor, we could almost triple the number of samples for analysis. Enrichment for tumors with a TPS ≥ 50 allowed us to further increase the statistical power towards this currently relevant cut-off. The participants used a pre-defined protocol for their platform for 22C3 LDT on VBMU or on LBO. In diagnostic practice, however, local optimization and validation of IHC protocols (LDT and assays) are a common practice in pathology laboratories, since local pre-analytical conditions and slight local variations in machine calibration can influence staining. It is possible that such local fine-tuning of the protocols would have positively influenced the concordance between laboratories in this study.

Taken together, there has been an impressive proliferation of harmonization and comparative studies during the past 2 years aiming to resolve the dilemma of PD-L1 testing. Several editorials and recent reviews provide guidance for pathologists through this thick jungle of partly conflicting data [20–22]. Overall, the ever-growing published evidence and our own data suggest that PD-L1 IHC using a standardized protocol for 22C3 antibody on VBMU is feasible and provides satisfactory results in most centers. The CE-marked SP263 assay is confirmed as a valid alternative to the 22C3 pharmDx on ASL48 to select patients for first-line monotherapy with pembrolizumab. 22C3 PD-L1 IHC on LBO using a standardized protocol shows major staining variability between centers, suggesting that PD-L1 IHC on LBO is more

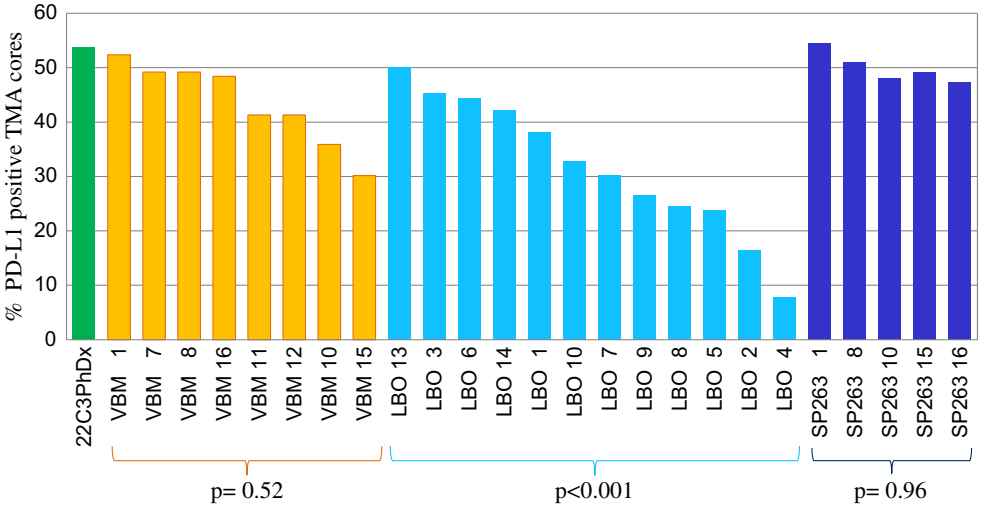

Fig. 4 Comparison of categorized PD-L1 results plotting the percentage of PD-L1-positive (TPS ≥ 50%) TMA cores by the DAKO 22C3phDx reference standard (green); the 22C3 VBM (yellow) in 8 centers; the 22C3 on LBO in 12 centers (light blue); and the SP263 in the 5 centers (dark blue). TPS tumor proportion score, VBM Ventana Benchmark, LBO Leica Bond

sensitive to local conditions. In any case, local fine-tuning of IHC protocols, rigorous local validation, and participation in ring trials remain critical when using a laboratory developed 22C3 IHC test. Monitoring of the local PD-L1 TPS rates for comparison with data from the literature and other laboratories is currently being used in Switzerland as another attractive tool for continuous quality control of PD-L1 IHC in routine diagnostic practice at www.biopath.ch [23].

Acknowledgments The authors acknowledge Petra Hirschmann and Betty Baschiera for establishing the 22C3 immunohistochemistry protocols.

Authors' contributions Spasenija Savic: Has substantially contributed to the conception of the work and the acquisition, analysis, and interpretation of data for the work. She has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally she has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Sabina Berezowska: Has substantially contributed to the acquisition and analysis of data for the work. She has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally she has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Serenella Eppenberger-Castori: Has substantially contributed to the analysis and interpretation of data for the work. She has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally she has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Gieri Cathomas: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Joachim Diebold: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Achim Fleischmann: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Wolfram Jochum: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Paul Komminoth: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Thomas McKee: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Igor Letovanec: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Zerina Jasarevic: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Matthias Rössle: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that

questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Gad Singer: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Michael von Gunten: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Andreas Zettl: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Roland Zweifel: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Alex Soltermann: Has substantially contributed to the acquisition of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Lukas Bubendorf: Has substantially contributed to the conception and design of the work and the acquisition, analysis, and interpretation of data for the work. He has revised the manuscript critically for important intellectual content and approved the version to be published. And, finally he has agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This study was funded by MSD Merck Sharp & Dohme AG.

Compliance with ethical standards

Conflict of interest Spasenija Savic attended advisory board meetings of MSD and Astra Zeneca, has a consulting role for MSD, and received speaker's honoraria from MSD and Roche.

Lukas Bubendorf attended advisory board meetings of Roche, MSD, Bristol-Myers Squibb and Astra Zeneca, and received financial research support from Roche and MSD. Sabina Berezowska and Joachim Diebold received consultancy honoraria from MSD. All remaining authors have declared no conflicts of interest.

References

1. Hirsch FR, McElhinny A, Stanforth D, Ranger-Moore J, Jansson M, Kulangara K, Richardson W, Towne P, Hanks D, Vennapusa B, Mistry A, Kalamegham R, Averbuch S, Novotny J, Rubin E, Emancipator K, McCaffery I, Williams JA, Walker J, Longshore J, Tsao MS, Kerr KM (2017) PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. J Thorac Oncol 12:208–222
2. Tsao MS, Kerr KM, Kockx M, Beasley MB, Borczuk AC, Botling J, Bubendorf L, Chirieac L, Chen G, Chou TY, Chung JH, Dacic S, Lantuejoul S, Mino-Kenudson M, Moreira AL, Nicholson AG, Noguchi M, Pelosi G, Poleri C, Russell PA, Sauter J, Thunnissen E, Wistuba I, Yu H, Wynes MW, Pintilie M, Yatabe Y, Hirsch FR (2018) PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project. J Thorac Oncol 13:1302–1311
3. Ratcliffe MJ, Sharpe A, Midha A, Barker C, Scott M, Scorer P, Al-Masri H, Rebelatto MC, Walker J (2017) Agreement between programmed cell death ligand-1 diagnostic assays across multiple protein expression cutoffs in non-small cell lung cancer. Clin Cancer Res 23:3585–3591
4. Adam J, Le Stang N, Rouquette I, Cazes A, Badoual C, Pinot-Roussel H, Tixier L, Danel C, Damiola F, Damotte D, Penault-Llorca F, Lantuejoul S (2018) Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann Oncol 29:953–958
5. Scheel AH, Baenfer G, Baretton G, Dietel M, Diezko R, Henkel T, Heukamp LC, Jasani B, Johrens K, Kirchner T, Lasitschka F, Petersen I, Reu S, Schildhaus HU, Schirmacher P, Schwamborn K, Sommer U, Stoss O, Tiemann M, Warth A, Weichert W, Wolf J, Buttner R, Ruschoff J (2018) Interlaboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. Histopathology 72:449–459
6. VENTANA PD-L1 (SP263) assay (CE IVD). In
7. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, Carcereny E, Ahn MJ, Felip E, Lee JS, Hellmann MD, Hamid O, Goldman JW, Soria JC, Dolled-Filhart M, Rutledge RZ, Zhang J, Lunceford JK, Rangwala R, Lubiniecki GM, Roach C, Emancipator K, Gandhi L, Investigators K (2015) Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 372:2018–2028
8. Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, Leiby MA, Lubiniecki GM, Shentu Y, Rangwala R, Brahmer JR, Investigators K (2016) Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 375:1823–1833
9. Bubendorf L, Lantuejoul S, Yatabe Y (2016) ALK and ROS1 analysis in cytology. In: Ed.: IASLC by editorial Rx press
10. VENTANA (2016) PD-L1 (SP263) assay staining of non-small cell lung cancer interpretation guide. Ventana Medical Systems, Inc., In
11. Neuman T, London M, Kania-Almog J, Litvin A, Zohar Y, Fridel L, Sandbank J, Barshak I, Vainer GW (2016) A harmonization study for the use of 22C3 PD-L1 immunohistochemical staining on Ventana's platform. J Thorac Oncol 11:1863–1868
12. Hendry S, Byrne DJ, Wright GM, Young RJ, Sturrock S, Cooper WA, Fox SB (2018) Comparison of four PD-L1 immunohistochemical assays in lung cancer. J Thorac Oncol 13:367–376
13. Ilie M, Khambata-Ford S, Copie-Bergman C, Huang L, Juco J, Hofman V, Hofman P (2017) Use of the 22C3 anti-PD-L1 antibody to determine PD-L1 expression in multiple automated immunohistochemistry platforms. PLoS One 12:e0183023
14. Roge R, Vyberg M, Nielsen S (2017) Accurate PD-L1 protocols for non-small cell lung cancer can be developed for automated staining platforms with clone 22C3. Appl Immunohistochem Mol Morphol 25:381–385
15. NordiQC: PD-L1 Assessment Run C4 2018. In
16. Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, Homer R, West WW, Wu H, Roden AC, Fujimoto J, Yu H, Anders R, Kowalewski A, Rivard C, Rehman J, Batenchuk C, Burns V, Hirsch FR, Wistuba II (2017) A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. JAMA Oncol 3:1051–1058

17. Marchetti A, Barberis M, Franco R, De Luca G, Pace MV, Staibano S, Volante M, Buttitta F, Guerini-Rocco E, Righi L, D'Antuono T, Scagliotti GV, Pinto C, De Rosa G, Papotti M (2017) Multicenter comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) assays to test PD-L1 expression for NSCLC patients to be treated with immune checkpoint inhibitors. J Thorac Oncol 12:1654–1663
18. Fujimoto D, Sato Y, Uehara K, Ishida K, Fukuoka J, Morimoto T, Kawachi H, Mori R, Ito M, Teraoka S, Nagata K, Nakagawa A, Otsuka K, Imai Y, Tomii K (2018) Predictive performance of four programmed cell death ligand 1 assay systems on nivolumab response in previously treated patients with non-small cell lung cancer. J Thorac Oncol 13:377–386
19. Munari E, Rossi G, Zamboni G, Lunardi G, Marconi M, Sommaggio M, Netto GJ, Hoque MO, Brunelli M, Martignoni G, Haffner MC, Moretta F, Pegoraro MC, Cavazza A, Samogin G, Furlan V, Mariotti FR, Vacca P, Moretta L, Bogina G (2018) PD-L1 assays 22C3 and SP263 are not interchangeable in non-small cell lung cancer when considering clinically relevant cutoffs: an Interclone evaluation by differently trained pathologists. Am J Surg Pathol 42:1384–1389
20. Kerr KM (2018) The PD-L1 immunohistochemistry biomarker: two steps forward, one step Back? J Thorac Oncol 13:291–294
21. Savic Prince S, Bubendorf L (2018) Predictive potential and need for standardization of PD-L1 immunohistochemistry. Virchows Arch
22. Scheel AH, Schafer SC (2018) Current PD-L1 immunohistochemistry for non-small cell lung cancer. J Thorac Dis 10:1217–1219
23. BIO-PATH: Biomarker tracker for pathologists. https://www.biopath.ch/mha/index.jsp