Assessment Run C3 2018

PD-L1

Amended version May 14th 2018§

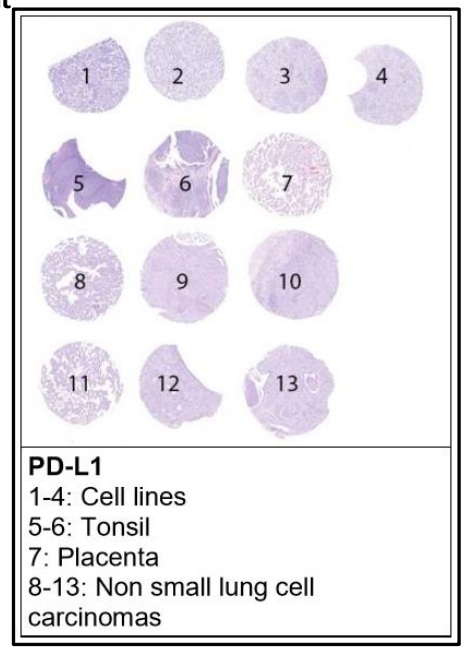
The third assessment in NordiQC Companion module C3 focused on the accuracy of the PD-L1 IHC assays performed by the participating laboratories to identify patients with non-small cell lung carcinoma (NSCLC) benefitting from immune therapy, either as first line treatment (Pembrolizumab [Keytruda®]) or second line treatment (Pembrolizumab, Nivolumab [Opdivo®])**§**. The PD-L1 expression levels in the circulated material used for the assessment were characterized by two CE IVD/FDA approved companion and complementary IHC assays, 22C3 pharmDX, SK006 Dako/Agilent, 28-8 pharmDX, SK005 Dako/Agilent, and the CE IVD approved SP263 (Ventana). Evaluation of the individual tissue cores and the associated cut-off values were used according to the interpretation guidelines provided by the manufacturer of these PD-L1 IHC assays.

§In the original version, Atezolizumab [Tecentriq®] was erroneously included.

Material

Table 1. **Content of the TMA used for the NordiQC PD-L1 C3 assessment**

	PD-L1 IHC TPS score*	**Eglible for treatment****
Cell line controls***		
1-4. Cell lines	NA	NA
Tissue controls		
5. Tonsil	NA	NA
6. Tonsil	NA	NA
7. Placenta	NA	NA
NSCLC		
8. NSCLC	No <1%	No/No
9. NSCLC	Low 1-49%	No/Yes
10. NSCLC	High ≥50%	Yes/Yes
11. NSCLC	No <1%	No/No
12. NSCLC	Low 1-49%	No/Yes
13. NSCLC	High ≥50%	Yes/Yes



* Tumour proportion score (TPS) determined by PD-L1 IHC 28-8, SK005 & 22C3, SK006 Dako performed in NordiQC reference lab.

** Using present recommendations for cut-off value of TPS ≥ 50% and 1-49% for first line treatment (Keytruda®) and second line (Keytruda® and Opdivo®), respectively.

*** Cell lines, HistoCyte (1-4). The series included a cell line with a negative TPS, very low TPS, intermediate/low TPS and high TPS. The cell lines were not included in the assessment but will later be used for digital image analysis.

All tissues were fixed in 10% neutral buffered formalin.
The participating laboratories were asked to perform the PD-L1 IHC assay, interpret the PD-L1 expression level and submit these scores to NordiQC. This allowed assessment of the technical performance (analytical accuracy) of the PD-L1 IHC assays and provided information on the reproducibility and concordance of the PD-L1 interpretation results among the laboratories.

PD-L1 IHC, Technical assessment

Criteria for assessing a staining as <u>Optimal</u> included:
The staining is considered perfect or close to perfect in all of the included tissues.
<u>TPS (as estimated by NordiQC assessor based on local staining) is concordant to the NordiQC reference data obtained in all 6 NSCLC cores.</u>

Criteria for assessing a staining as <u>Good</u> included:
The staining is considered acceptable in all of the included tissues. However, the protocol may be optimized to ensure the best staining intensity, counter staining, morphology and signal-to-noise ratio.
<u>TPS (as estimated by NordiQC assessor based on local staining) is still concordant to the NordiQC reference data in all 6 NSCLC cores.</u>

Criteria for assessing a staining as <u>Borderline</u> included:
The staining is considered insufficient, e.g., because of a generally too weak staining, a false negative staining or a false positive staining reaction of one of the included tissues. The protocol should be optimized.
<u>TPS (as estimated by NordiQC assessor based on local staining) is **not** found concordant to the NordiQC reference data in all 6 NSCLC cores.</u>

Criteria for assessing a staining as <u>Poor</u> included:
The staining is considered insufficient due to a false negative or a false positive staining reaction staining of more than one of the included tissues.
An optimization of the protocol is urgently needed.
<u>TPS (as estimated by NordiQC assessor based on local staining) is **not** found concordant to the NordiQC reference data in all 6 NSCLC cores.</u>

Participation

Number of laboratories registered for PD-L1 IHC C3	155
Number of laboratories returning PD-L1 IHC	146 (94%)
Number of laboratories returning PD-L1 scoring sheet	136 (88%)

Performance history
This was the third NordiQC assessment of PD-L1. Compared to the previous modules, an improved pass rate was obtained in C3 (see Table 2).

Table 2. **Proportion of sufficient results for PD-L1 in the three NordiQC runs performed**

	C1 2017	C2 2018	C3 2018
Participants, n=	68	145	146
Sufficient results	50%	84%	91%

Results: 146 laboratories participated in this assessment and 91% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Table 3 (page 3).

Table 3. **Assessment marks for IHC assays and antibodies run C3, PD-L1 IHC**

CE-IVD / FDA approved PD-L1 assays	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	Suff. OPS[2]
SP263, **790-4905**	52	Ventana/Roche	41	6	5	0	90%	92%
SP263, **790-4905**[3]	1	Ventana/Roche	1	0	0	0	-	-
22C3 pharmDX, **SK006**	27	Dako/Agilent	22	3	0	2	93%	100%
22C3 pharmDX, **SK006**[4]	8	Dako/Agilent	2	4	1	1	75%	-
28-8 pharmDX, **SK005**	5	Dako/Agilent	4	1	0	0	100%	100%
SP142, **740-4859**[5]	1	Ventana/Roche	0	0	0	1	-	-
Antibodies[6] for laboratory developed PD-L1 assays, conc. antibody		Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	Suff. OPS[2]
mAb clone **22C3**	32	Dako/Agilent	27	4	1	0	97%	100%
mAb clone **E1L3N**	6	Cell Signaling	3	3	0	0	100%	100%
mAb **CAL10**	2 3	Biocare Zytomed Systems	1	2	1	1	60%	100%
rmAb clone **28-8**	3	Abcam	3	0	0	0	-	-
rmAb clone **ZR3**	1 1	Cell Marque Zeta Corporation	2	0	0	0	-	-
rmAb clone **QR1**	2	Quartett / Biocyc	1	1	0	0	-	-
rmAb **BSR90**	1	Nordic Biosite	1	0	0	0	-	-
rmAb clone **SP142**[5]	1	Spring Biosystems	0	1	0	0	-	-
Total	146		108	25	8	5	-	-
Proportion			74%	17%	6%	3%	91%	-

1) Proportion of sufficient stains (optimal or good). 2) Proportion of sufficient stains with optimal protocol settings only, see below.
3) RTU system developed for the Ventana/Roche automated systems (BenchMark) but used by laboratories on a different platform (Leica Bond).
4) RTU system developed for the Agilent/Dako semi-automated systems (Autostainer Link48) but used by laboratories on different platforms (Ventana BenchMark and Dako Omnis).

5) assessed on TPS only, immune cells not included
6) mAb: mouse monoclonal antibody, rmAb: rabbit monoclonal antibody.

Detailed Analysis
CE IVD / FDA approved assays

SP263 (790-4905, Ventana): 41 of 52 (79%) protocols were assessed as optimal. Protocols with optimal results were typically based on heat induced epitope retrieval (HIER) in Cell Conditioning 1 (CC1), efficient heating time 52-64 min. on BenchMark Ultra, 16 min. incubation of the primary Ab and OptiView as detection kit. Using these protocol settings, 36 of 39 (92%) laboratories produced a sufficient staining result (optimal or good).

PD-L1 IHC 22C3 pharmDx (SK006, Dako): 22 of 27 (81%) protocols were assessed as optimal. Protocols with optimal results were based on HIER in EnVision™ Flex Target Retrieval Solution (TRS) low pH 6.1 (SK006) at 95-99°C for 20 min. in PT Link and 30 min. incubation of the primary Ab and EnVision Flex+ as the detection system on Autostainer Link 48. Using these protocol settings, 23 of 23 (100%) laboratories produced a sufficient staining result.

PD-L1 IHC 28-8 pharmDx (SK005, Dako): 4 of 5 (80%) protocols were assessed as optimal. Protocols with optimal results were based on HIER in EnVision™ Flex TRS low pH 6.1 at 97°C for 20 min. in PT Link and 30 min. incubation of the primary Ab and EnVison Flex+ as the detection system on Autostainer Link 48. Using these protocol settings, 5 of 5 (100%) laboratories produced a sufficient staining result.

Table 4 summarizes the proportion of sufficient and optimal marks for the most commonly used RTU systems. The performance was evaluated both as "true" plug-and-play systems performed strictly accordingly to the vendor recommendations and by laboratory modified systems changing basal protocol settings. Only protocols performed on the specific IHC stainer device are included.

Table 4. **Comparison of pass rates for vendor recommended and laboratory modified protocols**

CDx assay	**Vendor recommended protocol settings***		**Laboratory modified protocol settings****	
	Sufficient	Optimal	Sufficient	Optimal
Ventana BenchMark XT, GX, Ultra rmAb SP263, **790-4905**	23/26 (88%)	20/26 (77%)	25/26 (96%)	22/26 (85%)
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	23/23 (100%)	21/23 (91%)	2/4	1/4
Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005**	5/5 (100%)	4/5 (80%)	-	-

*Protocol settings recommended by vendor – Retrieval method and duration, Ab incubation times, detection kit, IHC stainer/equipment.
**Modifications in one or more of above mentioned parameters. Only protocols performed on the specified vendor IHC stainer are included.

Concentrated antibodies for laboratory developed (LD) assays

mAb **22C3**: 27 of 32 (84%) protocols were assessed as optimal. 17 optimal protocols were performed on BenchMark (Ventana) and 7 on Omnis (Dako).
On BenchMark GX/XT/Ultra, Ventana, the protocols providing an optimal result were typically based on a titre of 1:30-50, incubation time of 32-64 min., HIER in CC1 (efficient heating time 32-64 min.) and OptiView (8+8 min.) as detection system. Using these protocol settings, 8 of 8 (100%) laboratories produced a sufficient staining result.
7 laboratories used OptiView with Amplification kit and all provided an optimal staining result. Adding OptiView Amplification to above mentioned protocol settings 5 of 5 (100%) laboratories produced a sufficient staining result.
On Omnis, the protocols providing an optimal result were typically based on a titre of 1:20-50, incubation time of 30-40 min., HIER in TRS low pH 6.1 (Dako) at 97°C (efficient heating time 30-40 min.) and EnVision FLEX+ (10 min. in linker and 20-40 min. in polymer) as detection system.
Using these protocol settings, 7 of 7 (100%) laboratories produced a sufficient staining result.

mAb **E1L3N**: Three protocols provided an optimal result. All were based on HIER using an alkaline-buffer at 95-100°C for 30-40 min. The mAb clone E1L3N was diluted in the range of 1:100-400, incubated for 30-40 min. at room temp., and using a 3-layer technique as detection system. Using these protocol settings, 5 of 5 (100%) laboratories produced a sufficient staining result.

mAb **CAL10**: One protocol provided an optimal result. It was based on HIER using an alkaline buffer, diluted 1:50 with an incubation time for 30 min. at room temp. A 3-layer technique was used as detection system. Using these, protocol settings, 2 of 2 (100%) laboratories produced a sufficient staining result.

Table 5. **Optimal results for PD-L1 for the most commonly used antibodies as concentrates on the 3 main IHC systems***

Concentrated antibodies	**Ventana BenchMark GX/XT/Ultra**		**Dako Agilent Autostaininer/Omnis**		**Leica Bond III/Max**	
	CC1 pH 8.5	CC2 pH 6.1	TRS pH 9.0	TRS pH 6.1	BERS2 pH 9.0	BERS1 pH 6.0
mAb clone **22C3**	17/18** (94%)	-	-	9/10 (90%)	1/2	0/1
mAb clone **E1L3N**	0/1	-	1/1	-	2/3	0/1

*Antibody concentration applied as listed above, HIER buffers and detection kits used as provided by the vendors of the respective platforms.
**number of optimal results/number of laboratories using this buffer

Block construction and assessment challenges

The tissue micro array (TMA) blocks constructed for this PD-L1 run consisted of 4 cell lines, 6 NSCLCs, 2 tonsils and 1 placenta tissue. The NSCLCs were carefully selected so the slides cut from the blocks would contain 2 NSCLCs in each of the TPS groups: TPS negative (<1% PD-L1 positive tumour cells), TPS low (≥1-49%) and TPS high (≥50%). Reference slides throughout the block were stained using the approved pharmDx kits (SK006 and SK005, Dako/Agilent) and SP263 (790-4905, Ventana) for the assessment. During the assessment, a minor proportion of slides stained by the participants, revealed PD-L1 expression heterogeneity that was not seen in the initial screening of the tumours: NSCLC core no. 11, that initially had been scored as TPS negative, displayed small areas with distinct PD-L1 staining in more than 1% of the tumour cells. This pattern was also seen in the neighboring reference slides. In addition, both tissue core no. 9 and 12, initially scored as TPS low in the screening, were almost completely negative for PD-L1 (TPS negative). These heterogenic staining patterns were also seen in neighboring reference slides. Finally, and seen after staining in a fraction of the distributed slides, several of the cores (primarily tissue core no 9 and 12) were missing. The reason for this is uncertain, but it could be related to the tumour tissue itself, cutting of the sections or adherence to the glass slides.

Heterogeneity in PD-L1 expression is well known in NSCLCs and the assessment emulated clinical settings in this way. However, the inconsistent expression of PD-L1 in one or more of the included tissue cores was challenging for the assessment settings, defining protocols providing a sufficient result from protocols giving an insufficient result. During the assessment, slides with aberrant TPS category were compared to neighboring reference slides (stained by the NordiQC reference laboratory using approved kits). If the reference slides showed the same pattern, the TPS category was accepted without downgrading the assessment mark. The tissue cores 9 and 12 were the best indicators for assessing the overall sensitivity of the assay applied. For this reason, these cores were essential for evaluating the technical quality in this assessment. Slides that were missing tissue cores no 9 and 12 could only be assessed on the TPS high NSCLCs (tissue core no. 10 and 13), which in most cases was less challenging for the laboratories to stain. One might speculate that slides missing tissue cores no. 9 and 12 and receiving an optimal mark, would have been downgraded if tissue cores were intact and had been represented on the slide. However, it is important to note that no laboratory was downgraded based on missing cores or tissue heterogeneity.

Comments
In this third NordiQC run for PD-L1 in the companion module, C3, an overall pass rate of 91% was achieved. Insufficient PD-L1 IHC staining results were most frequently characterized by a reduced proportion of PD-L1 positive cells compared to the level expected as defined by the two PD-L1 IHC pharmDx assays, SK005 and SK006 (Dako/Agilent). This resulted in a too low TPS in on or more of NSCLC cores (9-10, 12-13).
The tissue cores 9 and 12 were the most challenging. Both cores were characterized as TPS low (≥1-49%) in the NordiQC reference labs and displayed a weak to moderate staining intensity in a small proportion of tumour cells. In insufficient results, one or both of these tumour cores were typically categorized as TPS negative (<1%) by the NordiQC assessor group. This could, in clinical settings, potentially prevent a patient from receiving second line treatment. This pattern was the most common (76%, 10 of 13) reason for insufficient staining. Importantly, both these tumours cores displayed heterogenic reaction patterns and all slides assessed were carefully examined and compared to neighboring reference slides to unravel any aberrant staining pattern that could be incorrectly interpreted and have an impact on the final assessment score (see explanation above).

The NSCLC tissue core 10 was characterized by the reference labs as TPS high and displayed a weak to moderate staining intensity in 90% of the tumour cells. A few labs (n=3) were challenged by this core,

since the tissue core 10 in the submitted slides was assessed as either TPS low or TPS negative by the NordiQC assessor group.

The NSCLC cores 8 and 11 were expected to be overall negative for PD-L1. As mentioned above, subsequent analysis revealed small areas of PD-L1 positivity in tissue core no. 11. Two laboratories produced results, showing relative large areas with PD-L1 staining reaction in both tissue core 8 and 11. These slides were carefully compared with neighboring reference slides displaying no PD-L1 reaction. The staining pattern for these two slides were therefore considered to be false positive staining. Finally, one laboratory experienced technical issues (related to inadequately levelled stainer rack on the Autostainer) that caused large false-negative areas.

The Ventana PD-L1 IHC assay 790-4905, based on the SP263 clone, was the most widely used assay for demonstration of PD-L1 and provided a pass rate of 89%. Applying protocol settings in compliance with the vendor recommendations the pass rate was 88% (23 of 26). In comparison, protocols based on laboratory modified protocol settings obtained a pass rate of 96% (25 of 26). One lab used the assay off label on a Leica Bond platform and obtained a sufficient result.

The Dako Agilent 22C3 pharmDx assay SK006 provided an overall pass rate of 89% (31/35). When applied according to the vendor recommendations, a pass rate of 100% was observed (23/23). 11 laboratories used the RTU product with other protocol settings than the recommended and obtained a significantly lower pass rate of 64% (7/11). Interestingly, 8 of these labs applied the RTU product on another stainer platform than the intended Dako Autostainer. One lab used the Dako Omnis and obtained an optimal result, while 7 labs used the Ventana BenchMark Ultra platform, providing a pass rate of 71% (5/7). As with the Ventana PD-L1 IHC assay (790-4905), it must be emphasized that off-label use of approved assays cannot be recommended as it require an extended (and sometimes challenging) internal validation.

The Dako/Agilent 28-8 pharmDx assay SK005 applying protocol settings in compliance with the vendor recommendations had an overall pass rate of 100% (5/5).

Grouped together, and using vendor recommended protocol settings, the three approved PD-L1 IHC assays, 22C3 SK006 Dako, 28-8 SK005 Dako and SP263 790-4905 Ventana gave a pass rate of 94% (51 of 54). Note, that SP263 is CE marked but not FDA approved in relation to NSCLC.

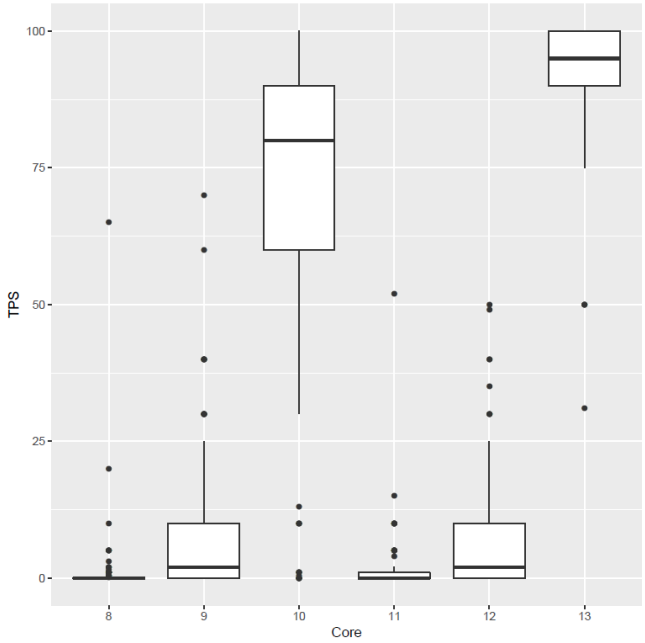
Laboratory developed (LD) assays (based on a concentrated Ab or a RTU product not used strictly to the recommendations of the developer) were used by 62% (91 of 146) of the participants. For this group a pass rate of 91% (83 of 91) was observed. As seen above, this is on the same level as approved PD-L1 IHC assays and is a significant improvement compared to run C2 providing a pass rate of 73%. This observation has to be interpreted within the limit of this assessment, considering the suboptimal quality of the material circulated.
The mAb clone 22C3 was the most widely used Ab within a LD assay (n=44) and the pass rate was 93% (41 of 44). This is a significant increase compared to the C2 run, where 76% of the LD assays based on this clone was sufficient. However, compared to corresponding pharmDx assay (SK006), which obtained a pass rate of 93% (81% optimal), LD assays can still be further improved. It is worth to note, that NordiQC has published an internally validated set of protocols for PD-L1 detection using the 22C3 as concentrated format on the major staining platforms (Røge R, Vyberg M, Nielsen S. Accurate PD-L1 Protocols for Non-Small Cell Lung Cancer can be Developed for Automated Staining Platforms With Clone 22C3. Appl Immunohistochem Mol Morphol. 2017 Jul;25(6):381-385.).

On the Ventana Benchmark platform, and using the mAb clone 22C3 within a LD assay, nine laboratories applied OptiView with amplification (tyramide based) as detection system and all were assessed as sufficient of which 89% were optimal. It is well known from previous assessments in NordiQC, both for other epitopes (general module) and PD-L1, that assays based on tyramide amplification can be challenging as low level expressing tissue structures may be negative and if not carefully calibrated, can cause false positive staining result. In general, tyramide amplification will enhance high level expressing cellular structures and may add a fine granular staining of structures expected to be negative. In this assessment, the majority of slides displayed the aberrant granular staining pattern of e.g. immune cells, but, this staining pattern was accepted as it did not significantly interfere with interpretation. However, since future scoring systems, as the CPS (combined positive score), also takes positive immune cells in account, this aberrant staining pattern will prevent correct scoring and should be avoided.

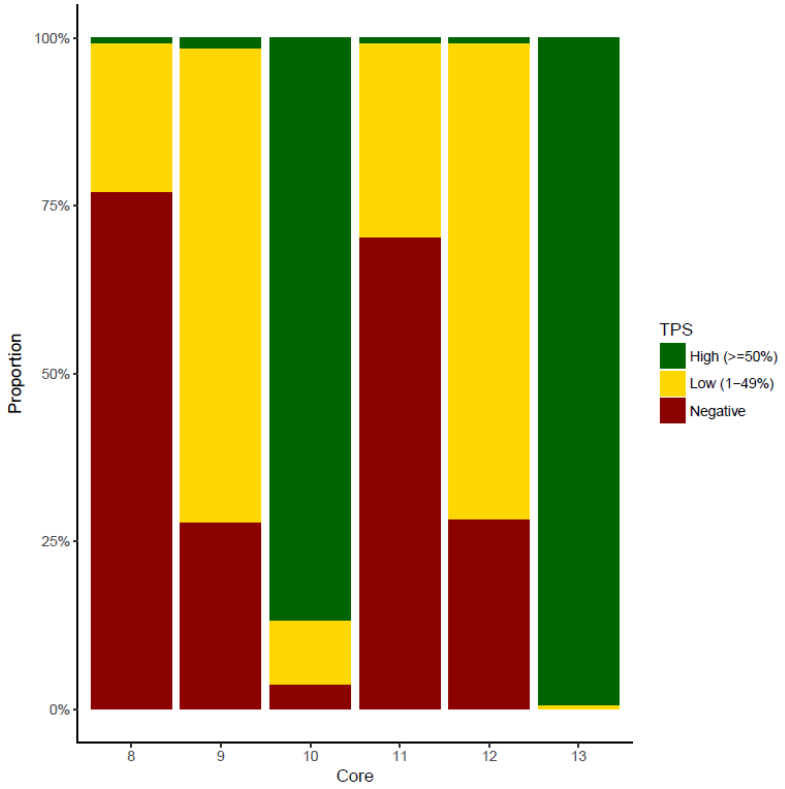
In this assessment several clones could be used to provide an optimal result: 22C3, 28-8, E1L3N, CAL10, ZR3, QR1 and BSR90. The companion diagnostic PD-L1 IHC assays from Dako/Agilent and Ventana/Roche

provided a high proportion of sufficient and optimal results. Within LD-assays, and no matter which Ab clone is used, meticulous calibration and validation of the assay is required.

PD-L1 interpretation and scoring consensus:
Participants were asked to evaluate the percentage of PD-L1 positive tumour cells in each of the six NSCLCs included in the assessment. The overall interpretation of PD-L1 expression and consensus rates of the participants are shown in Graph 1 and 2.

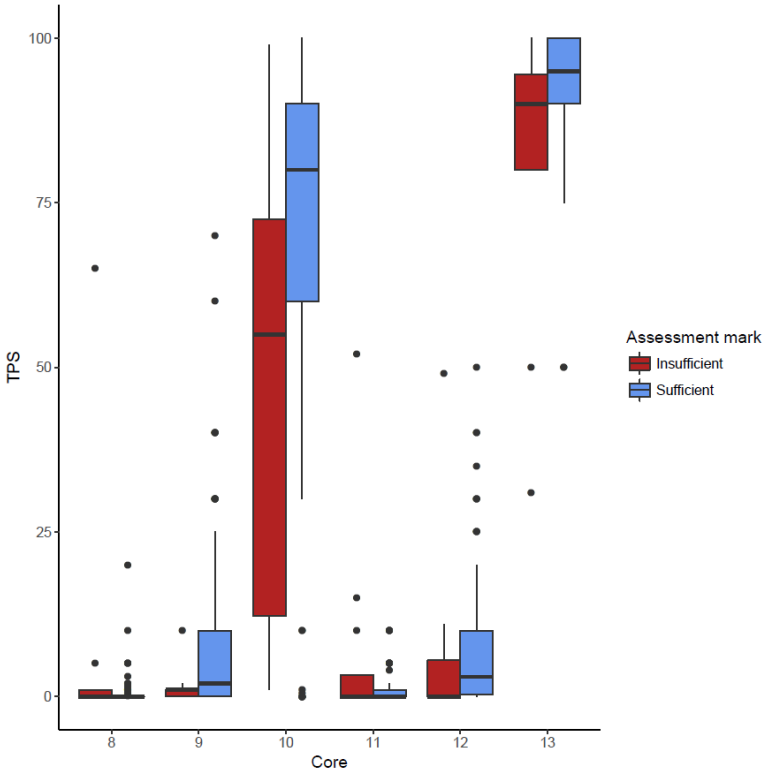


Graph 1. **NordiQC PD-L1 run C3: Participants' TPS scores (interpretation of the percentage of positive tumour cells)**



Graph 2. **NordiQC PD-L1 run C3: participant interpretation of PD-L1 TPS – impact on treatment**

As seen in Graph 1 and 2, relative high consensus rates for PD-L1 interpretation by the participants were observed. This was especially true for the tissue core 10 and 13. For the tissue core 8, 9, 11 and 12 about 25% had scored the cores in another TPS category than expected: For the tissue cores 8 and 11, that were expected to be TPS negative some participant had scored this as TPS low (≥1-49%). This could in some degree be explained by PD-L1 expression heterogeneity seen in core no 11. Both cores also contained PD-L1 positive macrophages that participants may have been interpreted as tumour cells. For the tumour cores 9 and 12, that were expected to be TPS low (≥1-49%), some participant scored this as TPS negative. This could also be explained by PD-L1 expression heterogeneity, and other slides completely missing these tumour cores due to less than optimal material circulated.
When stratifying for the assessment marks, analysis revealed that scores among participants that had received an insufficient mark (borderline or poor) reported lower TPS scores than laboratories that had received a sufficient mark, see Graph 3. However, this difference was not statistically significant and a marked overlap in scores between the two groups was seen. Additionally, laboratories that received an insufficient mark tended to have more diverse TPS scores.



Graph 3. **NordiQC PD-L1 run C3: interpretation concordance for labs with sufficient vs. insufficient results**

Controls

Tonsil and placenta were used as positive and negative tissue controls. In this assessment, tonsil was found to be superior to placenta, as tonsil displayed a range of PD-L1 expression levels. Using PD-L1 IHC 28-8, SK005/22C3, SK006 (Dako/Agilent) and SP263, 790-4905 (Ventana/Roche) and obtaining an optimal staining result, tonsil displayed the following reaction pattern: No staining reaction in the vast majority of lymphocytes including mantle zone and germinal centre B-cells, no staining reaction in superficial epithelial cells, a weak to moderate, typically punctuated membranous staining reaction of the majority of germinal centre macrophages and finally a moderate to strong staining reaction of the majority of epithelial crypt cells. SP263, 790-4905 Ventana/Roche provided similar staining pattern, but with an increased number of immune cells positive.

However, it was observed that a fully acceptable staining pattern in tonsil could be obtained together with insufficient and false negative result in the NSCLC. This underlines the need to identify more reliable positive tissue controls for PD-L1 and/or improve the interpretation criteria for a sufficient staining reaction in tonsil e.g. more accurately specify number and intensity of cells expected to be demonstrated.

Cell lines from HistoCyte (Newcastle, UK) were included in this assessment, primarily to evaluate if this material, in combination with digital image analysis, can be used to evaluate staining quality for PD-L1 and potentially be used as standard reference material for the validation of the precision of PD-L1 IHC assays. Subsequent analysis will be performed by NordiQC and published at a later stage.

Conclusion

This was the third NordiQC assessment of PD-L1 in the companion module. 146 laboratories participated and a pass rate of 91% was observed, which was an improvement compared to the previous run, C2. The three companion diagnostic PD-L1 IHC assays 28-8, SK005 Dako/Agilent, 22C3, SK006 Dako/Agilent and SP263, 790-4905 Ventana/Roche provided a pass rate of 91%. LD assays for PD-L1 provided a pass rate of 93%, comparable to the approved kits. However, the assessment of C3 was challenged by less than optimal material circulated (and for a minor fraction of the participant the included NSCLCs displayed varying degrees of PD-L1 expression heterogeneity). Additionally, some slides were missing critical cores. It must be underlined, that no lab was downgraded based on the quality of the circulated slides. However, this decision may have provided overall higher pass rate compared to if the circulated material had been of the required quality.

Tonsil is at present the preferred choice as positive and negative tissue control for PD-L1. The majority of epithelial crypt cells must show a moderate to strong staining reaction, while the germinal centre macrophages must display a weak to moderate membranous staining reaction. No staining must be seen

in the vast majority of lymphocytes and superficial epithelial cells.

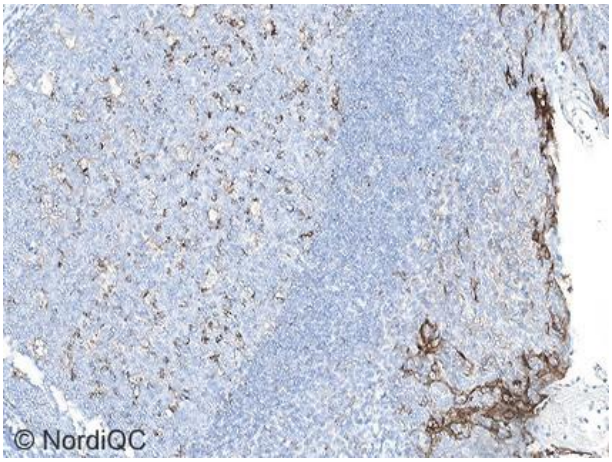


Fig. 1a. Optimal staining result of tonsil using the pharmDX IHC PD-L1 assay, SK006, Dako/Agilent based on the mAb clone 22C3. Same protocol used in Figs. 2a–4a. Crypt epithelial cells show a moderate to strong staining reaction, while the majority of germinal centre macrophages show a weak to moderate membranous staining reaction. The vast majority of lymphoid cells are negative.

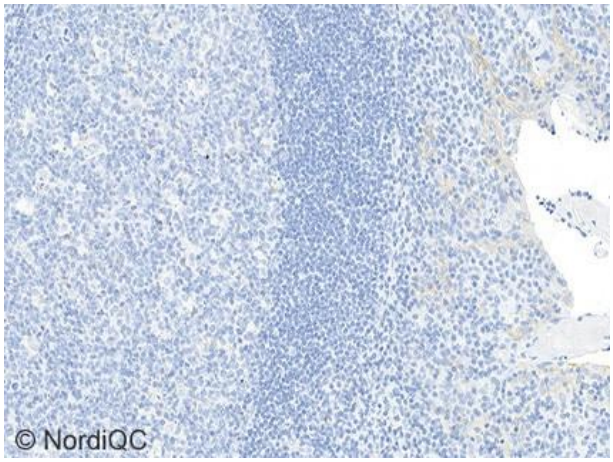


Fig. 1b. Insufficient staining result of tonsil using the pharmDX IHC PD-L1 assay, SK006, Dako/Agilent based on the mAb clone 22C3 with too short HIER and EnVision as detection system. Same protocol used in Figs. 2b-3b. Only a faint membranous staining reaction is seen in few crypt epithelial cells. Virtually all other cell types are negative. Compare with optimal staining in Fig.1a.

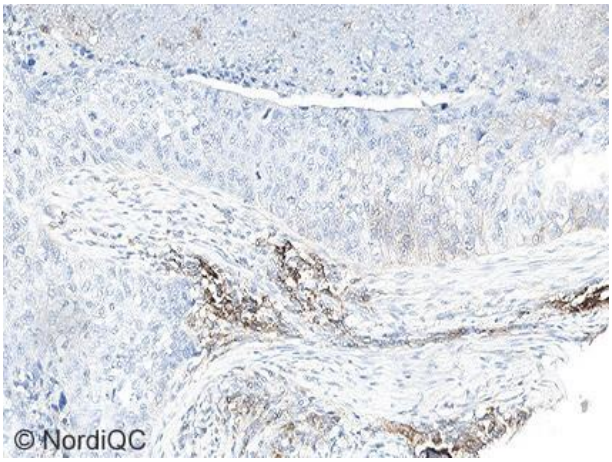


Fig. 2a. Optimal staining result of the tumour core no. 9 using same protocol as in Fig. 1a. Five percent of the neoplastic cells show a week to moderate and distinct membranous staining reaction. Also macrophages and dispersed lymphocytes are demonstrated. The tumour was categorized as TPS low (1-49%) and thus eligible for second line immune therapy. Same staining pattern was seen in reference slides.

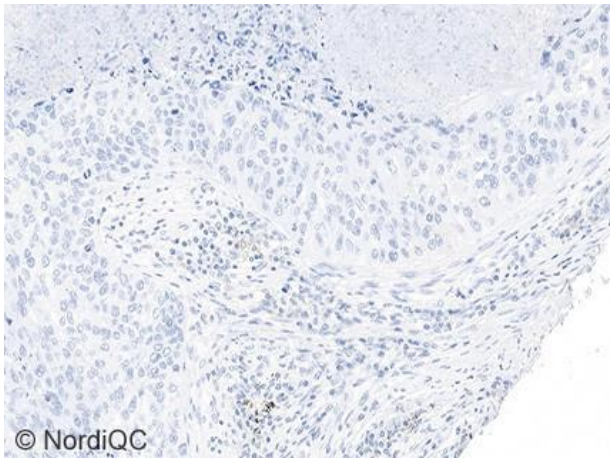
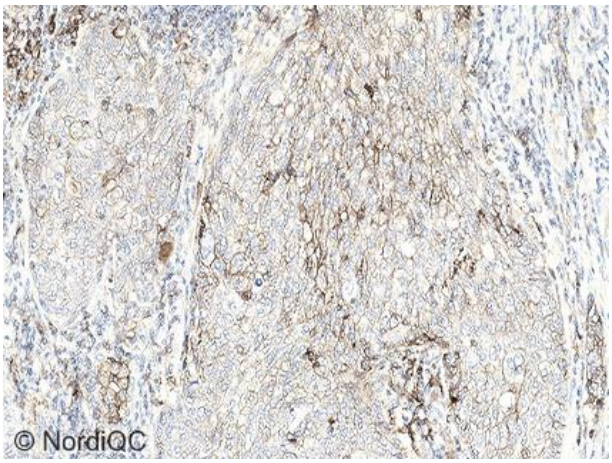


Fig. 2b. Insufficient staining result of the tumour core no. 9 using same protocol as in Fig. 1b. Virtually all tumour cells are negative providing a TPS below 1%. Also, macrophages and dispersed lymphocytes displays too weak staining intensity. Compare with optimal staining in Fig. 2a.



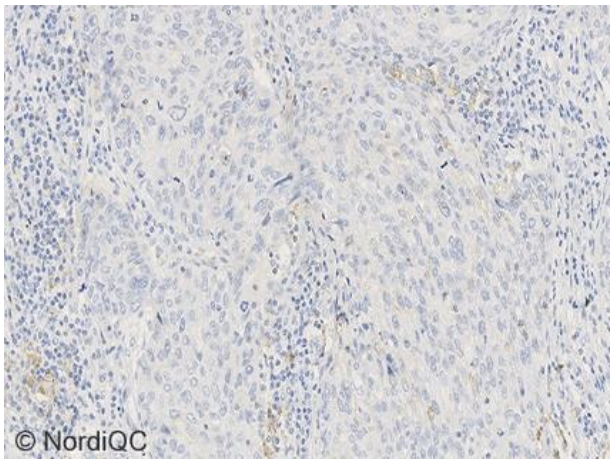


Fig. 3a. Optimal staining result of the tumour core no. 10 using same protocol as in Figs. 1a-2a.
The majority of neoplastic cells show a weak to moderate, distinct partial to complete membranous staining reaction. The tumour was categorized as TPS high (≥50%) and thus eligible for first line immune therapy.

Fig. 3b. Insufficient staining result of the tumour core no. 10 using same protocol as in Figs. 1b-2b.
The vast majority of tumour cells are negative for PD-L1 and TPS is significantly below 50%. Scattered immune cells only display a weak to moderate staining reaction. The assay has not been calibrated correctly, and does not provide the same PD-L1 staining pattern of tumour cells as obtained in Fig. 3a.

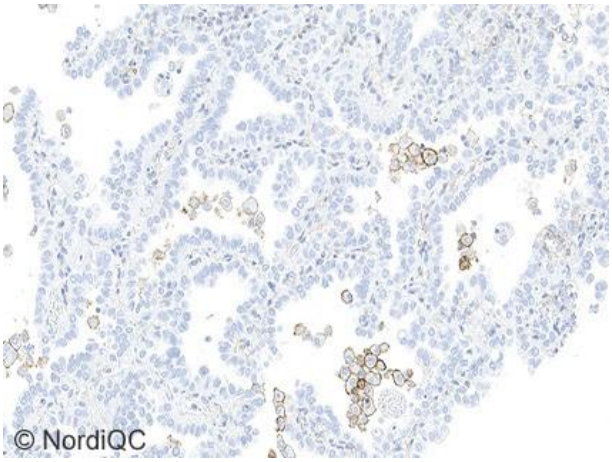


Fig. 4a. Optimal staining result of the tumour core no. 11 using same protocol as in Figs. 1a-3a.
The neoplastic cells are as expected negative for PD-L1. Only macrophages and dispersed lymphocytes show a distinct membranous staining reaction. The tumour was categorized as TPS negative (No <1%). Same staining pattern was observed in reference slides

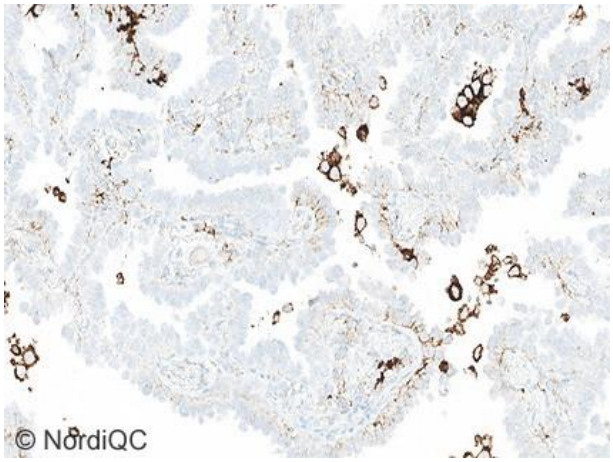


Fig. 4b. Insufficient staining result of the tumour core no. 11 using the pharmDX IHC PD-L1 assay, SK006, Dako/Agilent based on the mAb clone 22C3 but on a non-autostainer platform. The neoplastic cells, expected to be negative, show a granular membranous and false positive staining reaction. Compare with optimal result in Fig. 4a.

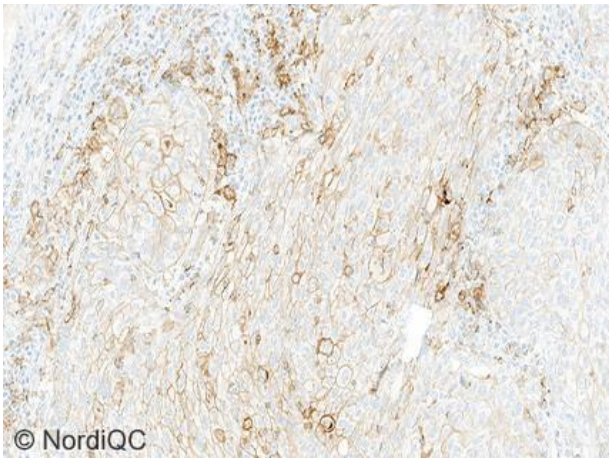


Fig. 5a. Optimal staining result of the tumour core no. 10 using the rmAb clone QR1 as a concentrate within a laboratory developed assay. The protocol was based on HIER in CC1 (Ventana/Roche) and UltraView as the detection system (Benchmark Ultra, Ventana/Roche). The majority of neoplastic cells show a weak to moderate, but distinct membranous staining reaction. The tumour was categorized as TPS high (≥50%).

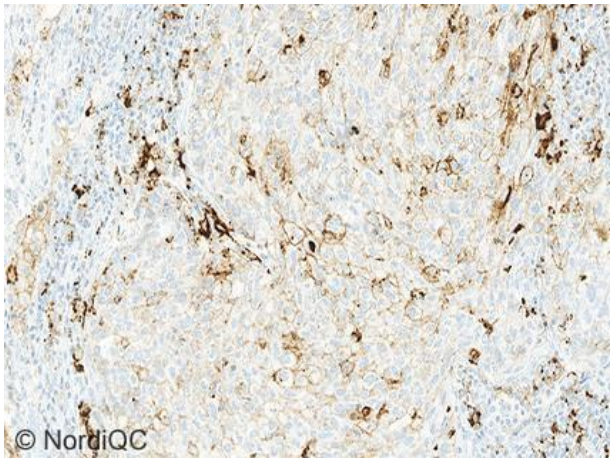


Fig. 5b. Staining for PD-L1 of the tumour core no. 10 using the mAb clone E1L3N with in a laboratory developed assay. Protocol was based on HIER in CC1 (Ventana/Roche) and OptiView with Amplification as detection system (BenchMark Ultra, Ventana/Roche). Although the tumour was categorized as TPS high, the majority of neoplastic cells displays a granular reaction pattern (due to the amplification step), rather than the continuous and homogenous staining pattern seen in Fig. 5a. If protocols are not carefully calibrated with this system, the interpretation of the reactions can be obscured due to granular deposit formed with tyramide based detection systems or the reactions can be too weak, risking that patients are positioned in wrong TPS

categories.

HLK/MB/MV/LE/RR 19.04.2018