Assessment Run C2 2018

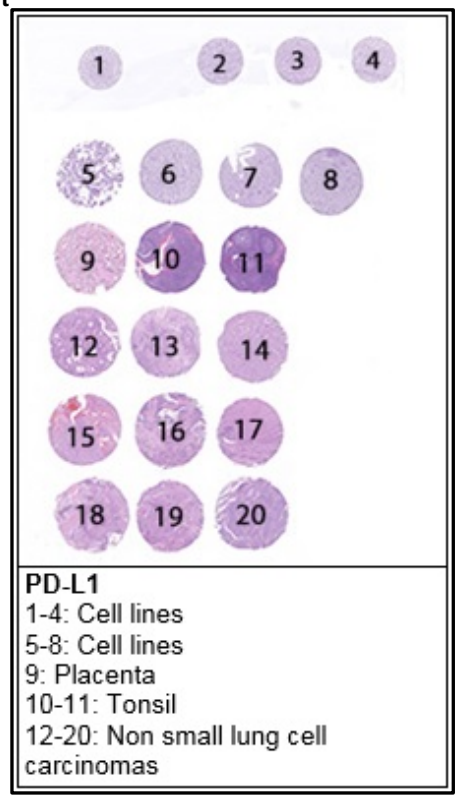
PD-L1

The second assessment in this new NordiQC Companion module C2 focused on the accuracy of the PD-L1 IHC assays performed by the participating laboratories to identify patients with nonsmall cell lung carcinoma (NSCLC) to be treated with immune therapy as either first line treatment (Keytruda®) or second line treatment (Keytruda® and Opdivo®). The PD-L1 expression levels in the circulated material used for the assessment were characterized by two CE IVD/FDA approved companion and commentary IHC assays, 28-8 pharmDX, SK005 Dako/Agilent and 22C3 pharmDX, SK006 Dako/Agilent. The associated cut-off values and interpretation guidelines were used accordingly to these two PD-L1 IHC assays.

Material

Table 1. **Content of the TMA used for the NordiQC PD-L1 C2 assessment**

	PD-L1 IHC TPS score*	Eglible for treatment**
Cell line controls* **		
1. Cell line 1	NA	NA
2. Cell line 2	NA	NA
3. Cell line 3	NA	NA
4. Cell line 4	NA	NA
5. Cell line 5	NA	NA
6. Cell line 6	NA	NA
7. Cell line 7	NA	NA
8. Cell line 8	NA	NA
Tissue controls		
9. Placenta	NA	NA
10. Tonsil	NA	NA
11. Tonsil	NA	NA
NSCLC		
12. NSCLC	No <1%	No
13. NSCLC	No <1%	No
14. NSCLC	No <1%	No
15. NSCLC	Excluded	Excluded
16. NSCLC	Low 1-49%	Yes
17. NSCLC	High ≥50%	Yes
18. NSCLC	High ≥50%	Yes
19. NSCLC	High ≥50%	Yes
20. NSCLC	High ≥50%	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 of 1-49% and ≥ 50% for second line (Keytruda® and Opdivo®) and first line treatment (Keytruda®), respectively.
*** Cell lines, Horizon Discovery, prod. Id HD788 (1-4). and Cell lines, HistoCyte (5-8). Both series included cell lines with a negative TPS, very low TPS, intermediate/low TPS and high TPS.

All tissues were fixed in 10% neutral buffered formalin.
Core no. 15 was excluded from the assessment due to heterogeneous expression and TPS throughout the block.
The participating laboratories were asked to perform the PD-L1 IHC assay accordingly to the protocol used in the laboratory, interpret the PD-L1 expression level and submit these scores to NordiQC. This allowed both assessment of the technical performance (analytical accuracy) of the PD-L1 IHC assays and information on the reproducibility and concordance of the interpretation of PD-L1 expression among the laboratories.

PD-L1 IHC, Technical assessment

Criteria for assessing a staining as Optimal included:
The staining is considered perfect or close to perfect in all of the included tissues.
TPS is concordant to the NordiQC reference data is obtained in all 8 NSCLC cores.

Criteria for assessing a staining as Good 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.
TPS is still concordant to the NordiQC reference data in all 8 NSCLC cores.

Criteria for assessing a staining as Borderline 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.
TPS is **not** found concordant to the NordiQC reference data in all 8 NSCLC cores.

Criteria for assessing a staining as Poor included:
The staining is considered very insufficient e.g., because of a false negative or a false positive staining reaction staining of more of the included tissues.
An optimization of the protocol is urgently needed.
TPS is **not** found concordant to the NordiQC reference data in all 8 NSCLC cores.

PD-L1 IHC, Interpretation
All participating laboratories were asked to submit a scoring sheet with their interpretation of the tumour proportion score (TPS) in the NSCLCs. Results were compared to NordiQC data from the reference laboratory to analyze scoring consensus.

Participation

Number of laboratories registered for PD-L1 IHC C1	147
Number of laboratories returning PD-L1 IHC	145 (99%)
Number of laboratories returning PD-L1 scoring sheet	126 (87%)

Performance history
This was the second NordiQC assessment of PD-L1 in NordiQC.
Compared to C1, an improved pass rate was obtained in C2 (see Table 2).

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

	Run C1 2017	Run C2 2018
Participants, n=	68	145
Sufficient results	50%	84%

Results: 145 laboratories participated in this assessment and 84% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Table 3 (page 3).

Detailed Analysis

CE IVD / FDA approved assays

PD-L1 IHC 22C3 pharmDx (SK006, Dako): 15 of 25 (60%) protocols were assessed as optimal. Protocols with optimal results were based on heat induced epitope retrieval (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, linker and polymer. Using these protocol settings, 22 of 23 (96%) laboratories produced a sufficient staining result (optimal or good).

PD-L1 IHC 28-8 pharmDx (SK005, Dako): 6 of 6 (100%) protocols were assessed as optimal. Protocols with optimal results were based on HIER in EnVision™ Flex TRS low pH 6.1 (SK005) at 97°C for 20 min. in PT Link and 30 min. incubation of the primary Ab, linker and polymer. Using these protocol settings, 6 of 6 (100%) laboratories produced an optimal staining result.

Table 3. **Assessment marks for IHC assays and antibodies run C2, PD-L1 IHC**

CE-IVD / FDA approved PD-L1 assays	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	Suff. OPS[2]
22C3 pharmDX, **SK006**	23	Dako/Agilent	15	7	0	1	96%	96%
22C3 pharmDX, **SK006**[4]	5	Dako/Agilent	1	2	0	2	60%	-
28-8 pharmDX, **SK005**	6	Dako/Agilent	6	0	0	0	100%	100%
SP263, 790-4905	49	Ventana/Roche	44	2	2	1	94%	98%
SP263, 790-4905[5]	2	Ventana/Roche	0	0	2	0	-	-
Antibodies[3] for laboratory developed PD-L1 assays, conc. antibody	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	Suff. OPS[2]
mAb clone **22C3**	39	Dako/Agilent	12	18	4	5	76%	-
mAb clone **E1L3N**	9	Cell Signaling	2	6	1	0	89%	-
mAb **CAL10**	2	Biocare	0	1	0	1	-	-
mAb **CAL10**	1	Zytomed	0	0	0	1	-	-
rmAb clone **28-8**	3	Abcam	1	1	1	0	-	-
rmAb clone **ZR3**	1	Zeta Corporation	1	0	0	0	-	-
rmAb clone **ZR3**	1	Nordic Biosite	1	0	0	0	-	-
rmAb clone **ZR3**	1	Gene Tech	0	1	0	0	-	-
rmAb clone **SP142**	1	Spring Biosystems	0	0	1	0	-	-
rmAb clone **QR1**	1	Quartett	1	0	0	0	-	-
rmAb clone **HDX3**	1	Halioseek	1	0	0	0	-	-
Total	145		85	38	11	11	-	-
Proportion			59%	26%	8%	8%	85%	-

1) Proportion of sufficient stains (optimal or good).
2) Proportion of sufficient stains with optimal protocol settings only, see below.
3) mAb: mouse monoclonal antibody, rmAb: rabbit monoclonal antibody.
4) RTU system developed for the Agilent/Dako`s semi-automated systems (Autostainer Link48) but used by laboratories on different platforms (Ventana Benchmark and Dako Omnis).
5) RTU system developed for the Ventana/Roche`s automated systems (BenchMark) but used by laboratories on different platforms (Leica Bond and Dako AS48).

SP263 (790-4905, Ventana): 44 of 51 (86%) protocols were assessed as optimal. Protocols with optimal results were based on HIER in Cell Conditioning 1 (CC1), efficient heating time 56-64 min. in BenchMark Ultra, 16-20 min. incubation of the primary Ab and OptiView as detection kit. Using these protocol settings, 46 of 49 (94%) laboratories produced a sufficient staining result.

Concentrated antibodies for laboratory developed (LD) assays

mAb **22C3**: 12 of 39 protocols were assessed as optimal result. 6 protocols were performed on BenchMark, Ventana, 4 on Omnis, Dako.
On BenchMark XT/Ultra, Ventana, the protocols providing an optimal result typically were based on a titre of 1:40, incubation time of 56-64 min., heat induced epitope retrieval (HIER) in Cell Conditioning 1 (CC1) (efficient heating time 48 min.) and OptiView (8+8 min in linker and multimer) as detection system.
Using these protocol settings, 7 of 7 (100%) laboratories produced a sufficient staining result.
On Omnis, the protocols providing an optimal result typically was based on a titre of 1:20, incubation time of 30-40 min., HIER in Target Retrieval Solution (TRS) low pH 6.1 (Dako) at 97°C (efficient heating time 40 min.) and EnVision FLEX+ (10 min. in linker and 30-40 min in polymer) as detection system.
Using these protocol settings, 5 of 5 (100%) laboratories produced a sufficient staining result.

mAb **E1L3N**: Two protocols provided an optimal result. One was based on HIER using Epitope retrieval solution 2 pH 9 (Bond, Leica) at 100°C for 30 min. on-board, Bond III (Leica) The mAb clone E1L3N was diluted 1:400, incubated for 30 min. at room temp. Refine (Leica) was used as detection system with an incubation time of 8 min. in linker and 8 min. in polymer. The protocol was performed on Bond III, Leica.
The other protocol was based on HIER in PT-module using Tris-EDTA/EGTA pH 9 at 98°C for 20 min. The mAb clone E1L3N was diluted 1:800, incubated for 30 min. at room temp. BrigthVision plus (Immunologic) was used as detection system with an incubation time of 30 min. in polymer. The protocol was performed on Autostainer, Thermo.

Comments

In this second NordiQC run for PD-L1 in the companion module C2, a pass rate of 84% was observed. Insufficient PD-L1 IHC staining results were most frequently characterized by a reduced proportion of cells being PD-L1 positive compared to the level expected as defined by the two PD-L1 IHC pharmDx assays, SK005 and SK006, Dako/Agilent. Consequently, a too low TPS in one or more of the NSCLCs was observed. The tissue cores 16-18 were the most challenging. The tumours in core 17 and 18 were both expected to be TPS high (≥50%), but only demonstrated a weak to moderate staining intensity (as characterized by SK005 and SK006 in the NordiQC reference laboratory). In the insufficient results, these two tumours typically were categorized as TPS Low (≥1-49%) that would move the immune therapy option from 1' line to 2' line. This pattern was seen in most of the insufficient cases.

The tumour in core 16 was expected to be TPS low (≥1-49%). In insufficient stains this tumour was either characterized as TPS negative or TPS high (≥50%), which also would impact the treatment options. 12 laboratories produced an insufficient result in core 16: In 7 cases it was characterized by a false negative result (TPS Neg), in 5 cases by a false positive result (TPS high).

The remaining insufficient results were characterized by poor signal-to-noise ratio and/or impaired morphology.

The Ventana PD-L1 IHC assay 790-4905, SP263 was the most widely used assay for PD-L1 and provided an overall pass rate of 98%, when based on protocol settings in compliance with the vendor recommendations. Two laboratories used the kit off-label on other staining platform (Leica Bond and Dako AS48) and they both produced an insufficient result.

The Dako/Agilent 22C3 pharmDx assay SK006 provided an overall pass rate of 96%, when applied by protocol settings in compliance with the vendor recommendations. However, a slightly reduced number of optimal results was observed, compared to the two other CE-IVD approved PD-L1 IHC assays (SK005, Dako/Agilent and SP263, Ventana/Roche) as shown in table 2.
Five laboratories used the kit off-label on another staining platform (BenchMark, Ventana or Omnis, Dako). When pooled, a pass rate of 60% was observed for these. It must be emphasized that off-label use of approved assays is problematic since the intended use is violated. It therefore requires an extended (and sometimes) challenging internal validation similar to that of laboratory developed (LD) assays.

The Dako/Agilent 28-8 pharmDx assay SK005, applied by protocol settings in compliance with the vendor recommendations, was most successful with an overall pass rate of 100%.

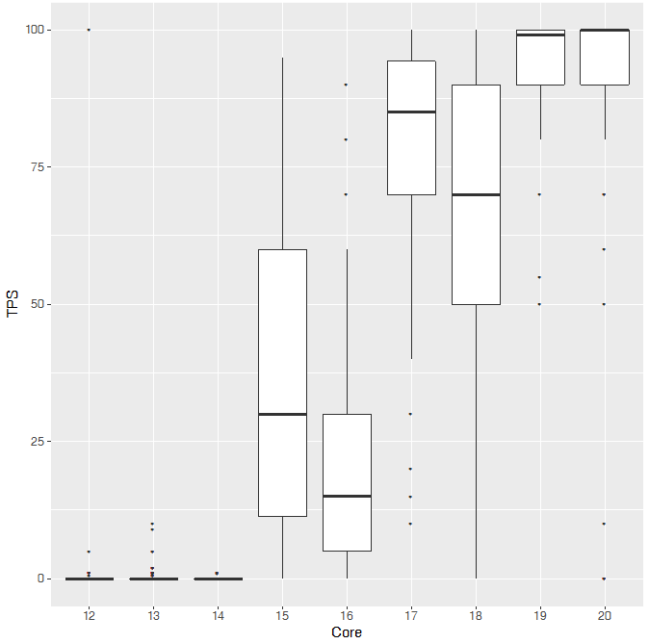
Grouped together, the three approved PD-L1 IHC assays, 22C3 SK006 Dako, 28-8 SK005 Dako and SP263 790-4905 Ventana gave a pass rate of 95% (74 of 78 protocols performed with the intended protocol settings). Note, that SP263 is CE marked but not FDA approved in relation to NSCLC.
Laboratory developed (LD) assays were used by 46% (67 of 145) of the participants. For this group a pass rate of 73% (49 of 67) was observed. The mAb clone 22C3 was the most widely used Ab within a LD assay (n=39). A reduced pass rate of 76% (compared to the pass rate of 96% for the corresponding pharmDx assay SK006) was seen. The reduced pass rate for LD assays compared to approved and vendor validated assays for PD-L1 IHC clearly indicates the challenge for the laboratories to identify and validate LD protocols to give results concordant to the vendor validated assays. However, compared to the previous run C1 for PD-L1, a marked improvement for LD assays based on mAb clone 22C3 was seen. As described in the detailed analysis for protocol settings providing optimal results for concentrated Abs, very successful protocols for PD-L1 using mAb clone 22C3 on Ventana BenchMark and Dako Omnis were identified. Caution must be taken due to limited number of observations, but a 100% pass rate (n=12) was seen when using "best practice" protocol settings as described.*

*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.

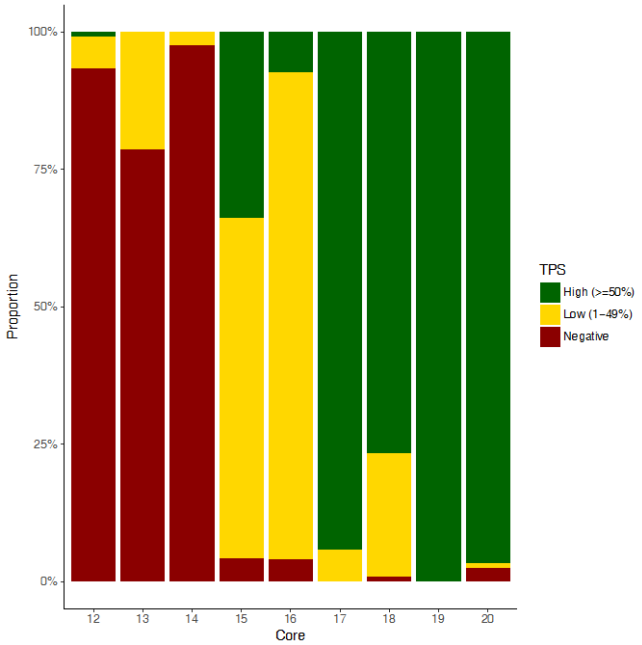
As seen in C1 for LD assays performed on BenchMark, Ventana, the use of amplification kit in combination with OptiView as detection system seemed to be less successful for mAb clone 22C3. From the experience of other NordiQC assessments, this tyramide based IHC system seems to give a binary staining result (either negative or strongly positive). This has shown to be very successful for ALK demonstration in lung adenocarcinoma, but seems to be less adequate to demonstrate an accurate and dynamic range of expression levels which is essential for PD-L1 interpretation and establishment of TPS using present guidelines. In this run, 10 protocols for mAb clone 22C3 applied on BenchMark were used with OptiView + amplification. An overall pass rate of 60% was seen and 10% optimal, which was a significantly inferior compared to the LD assay described in the detailed analysis.

As this was the second assessment for PD-L1 IHC only, data to identify reliable and recommended LD assays still are limited and conclusions must be drawn with caution. Several clones as 22C3, E1L3N, HDX3, QR1 and ZR3 could all be used to set-up protocols giving optimal staining results provided that a meticulous calibration and validation of the LD assays have been completed. However, this is no guarantee for similar successful results should other cut-off values be defined or other tumour types analysed.

PD-L1 interpretation and scoring consensus:
Participants were asked to evaluate the percentage of PD-L1 positive tumour cells in each of the nine 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.

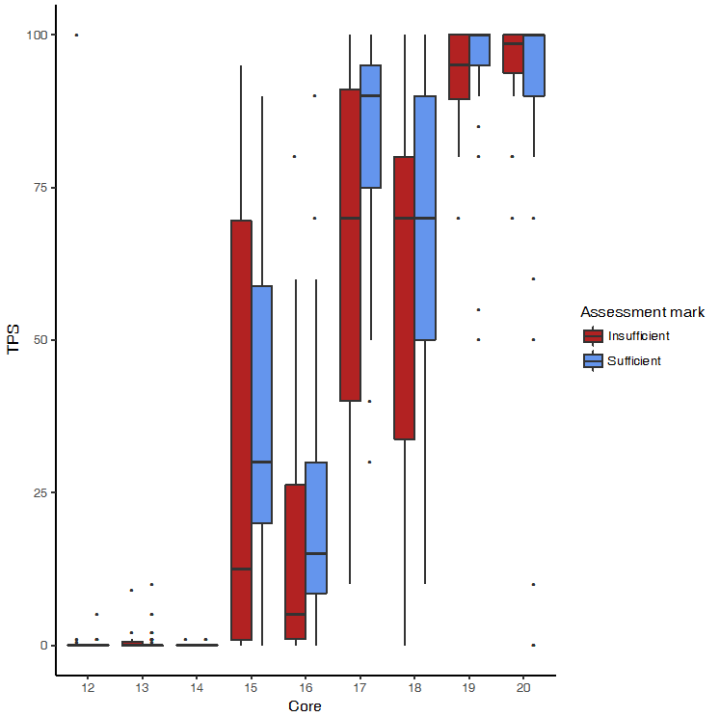


Graph 1. **NordiQC PD-L1 run C2: Participants' TPS scores (interpretation of the percentage of positive tumour cells).**



Graph 2. **NordiQC PD-L1 run C2: 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 a significant improvement compared to the last run (C1), where a high degree of disagreement was seen. The NSCLC tissue core no. 15 was excluded from the assessment due to heterogeneous PD-L1 expression. The submitted scoring sheet confirmed this, as 62% (n=73) categorized the tumour as low, while 34% (n=40) scored as high.

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 statistical significant and showed marked overlap in scores between the two groups. Additionally, laboratories that received an insufficient mark tended to have more diverse TPS scores.
The most challenging tissue cores were no 13 and 18. Among the returned scoring sheets, 21% scored tissue core no. 13 low. This tumour was consistently TPS neg. (both by NordiQC reference protocols and participants' assays), but included macrophages, which was interpreted as tumour cells by many participants. Concerning tissue core no. 18, 23% scored this as low. No explanation for this deviation could be extrapolated from the submitted slides.



Graph 3. **NordiQC PD-L1 run C2: 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, whereas placenta showed a more binary expression level with cells being either negative or strongly positive. 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, 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.
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 scenario was observed by the using Optiview + amplification within a LD assay typically based on mAb 22C3. 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 Horizon Discovery (Cambridge UK) and 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 second NordiQC assessment of PD-L1 in the new companion module. 145 laboratories participated and a pass rate of 84% was observed, which was a significant improvement compared to the previous run. The three companion diagnostic PD-L1 IHC assays 28-8, SK005 Dako/Agilent, 22C3, SK006

Dako/Agilent and SP263, 790-4905 Ventana/Roche were most successful providing a pass rate of 95%. LD assays for PD-L1 provided a pass rate of 73%, which was very encouraging compared to C1, in which a pass rate of 20% for this group was observed. Access to best practice protocol settings for LD assays are being developed and published supporting the implementation of LD assays. However, it must be emphasized that the LD assays must be carefully validated before use for diagnostics.

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 a weak to moderate membranous staining reaction. No staining must be seen in the vast majority of lymphocytes.

MV/RR 25.01.18 (30.01.18)

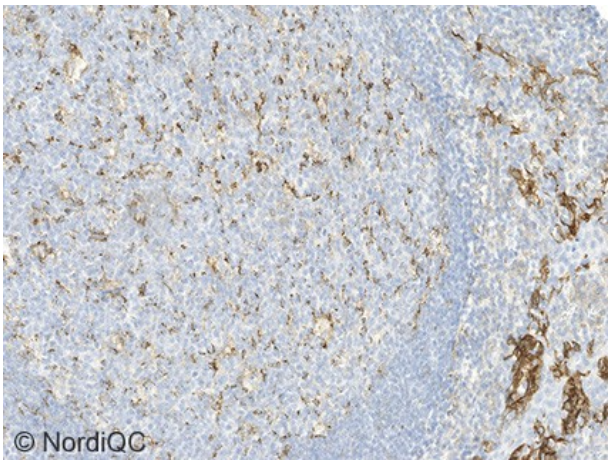


Fig. 1a. Optimal PD-L1 staining result of tonsil using the pharmDx IHC PD-L1 assay, SK006, Dako/Agilent based on the mAb clone 22C3. Crypt epithelial cells show a moderate to strong, staining reaction, while the majority of germinal centre macrophages show a weak to moderate membranous staining reaction. The vast majority of lymphoid cells are negative. Also compare with Figs. 2a–5a, same protocol.

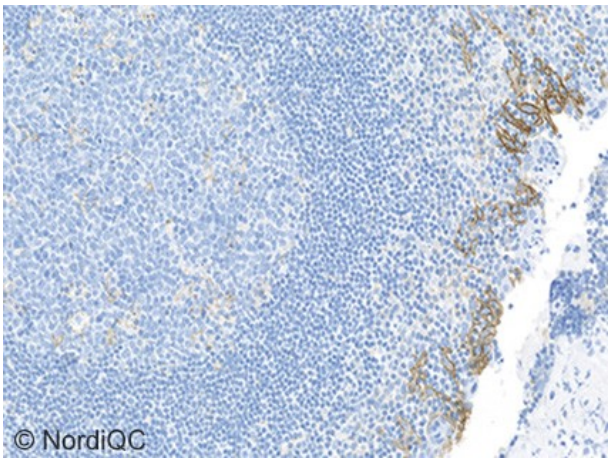


Fig. 1b. PD-L1 staining result of tonsil using the mAb clone 22C3 as concentrated format (M3653, Dako/Agilent) within a laboratory developed assay. The assay was based on OptiView + amplification and performed on BenchMark Ultra (Ventana/Roche). A moderate to strong staining reaction is seen in both crypt epithelial cells whereas in germinal centre macrophages virtually are negative. Also compare with Figs. 2b–5b, same protocol.

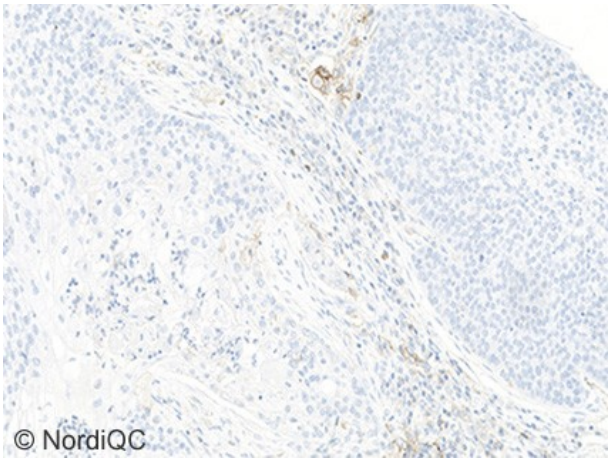


Fig. 2a. Optimal PD-L1 staining result of the NSCLC tissue core no. 13 using same protocol as in Fig. 1a. The neoplastic cells are all negative and only macrophages and dispersed lymphocytes show a distinct membranous staining reaction serving as internal positive tissue control. The tumour is categorized as TPS negative.

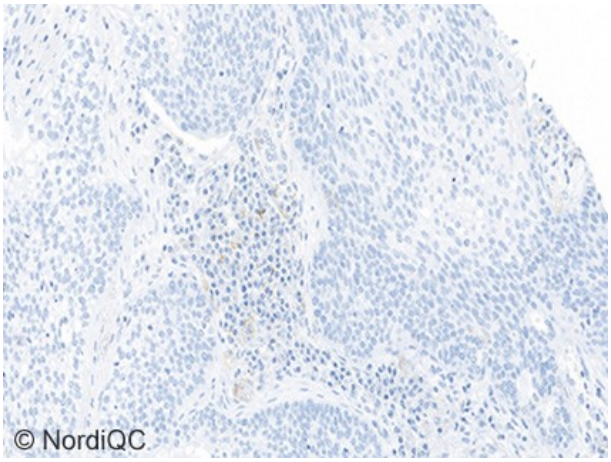


Fig. 2b. PD-L1 staining result of the NSCLC tissue core no. 13 using same protocol as in Fig. 1b. The neoplastic cells are all negative and only macrophages and dispersed lymphocytes are demonstrated. The tumour is categorized as TPS negative. However also compare with Figs. 3b–5b.

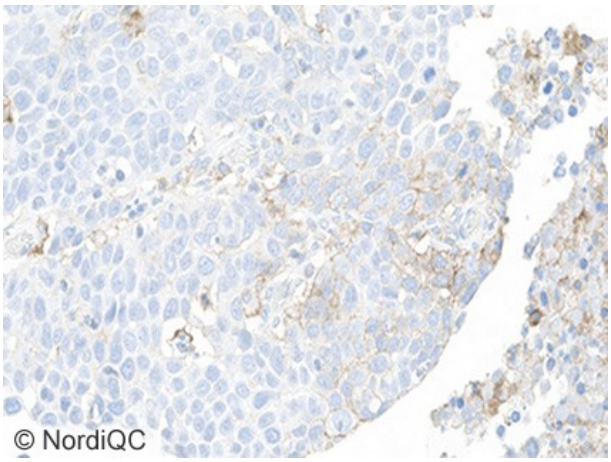


Fig. 3a. Optimal PD-L1 staining result of the NSCLC tissue core no. 16 using same protocol as in Figs. 1a and 2a. About 20-30% of the neoplastic cells show a week to moderate and distinct membranous staining reaction. The tumour is categorized as TPS low.

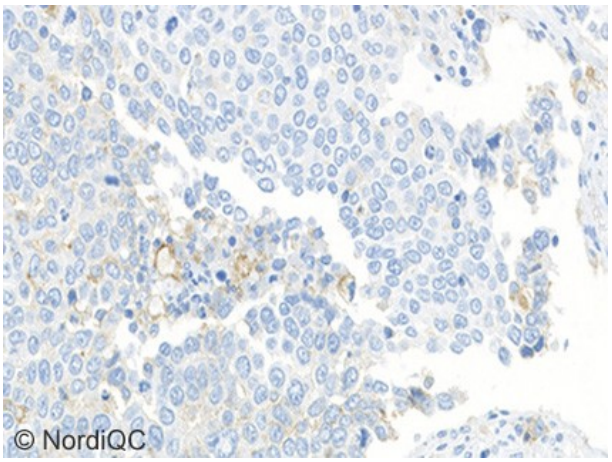


Fig. 3b. PD-L1 staining result of the NSCLC tissue core no. 16 using same protocol as in Figs. 1b and 2b. The neoplastic cells show a weak, granular membrane staining reaction, compared to the pattern seen in Fig. 3a.

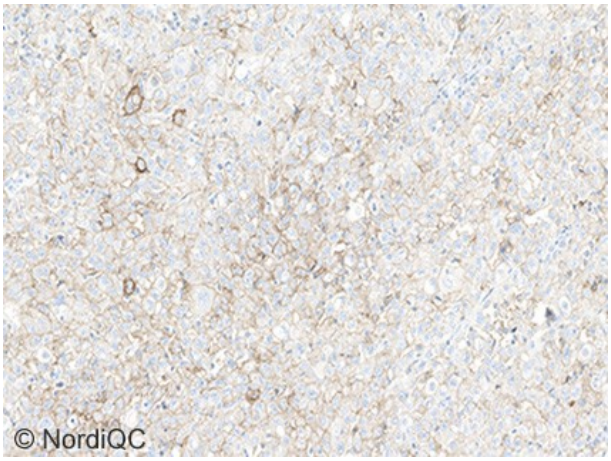


Fig. 4a. Optimal PD-L1 staining result of the NSCLC tissue core no. 17 using same protocol as in Figs. 1a-3a. The majority of neoplastic cells show a weak to moderate membranous staining reaction, in some areas only very weak and partial membranous staining reaction. The tumour is categorized as TPS high and thus eligible for 1' line immune therapy.

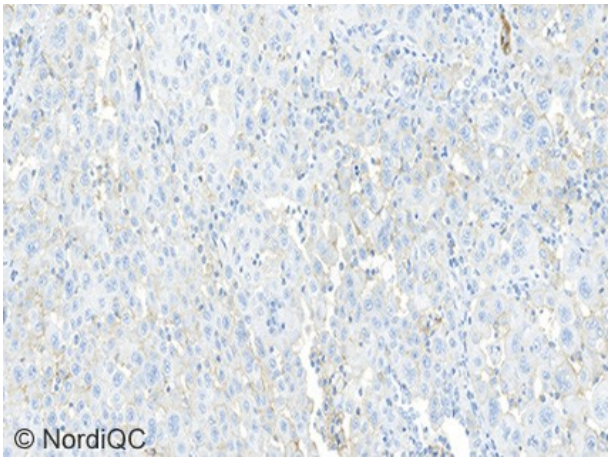


Fig. 4b. Insufficient PD-L1 staining result of the NSCLC tissue core no. 17 using same protocol as in Figs. 1b-3b. Only few cells show a distinct membranous staining reaction, whereas the vast majority are negative. Few immune cells show a strong staining reaction. The assay has not been calibrated to demonstrate the weak PD-L1 expression in the tumour cells.

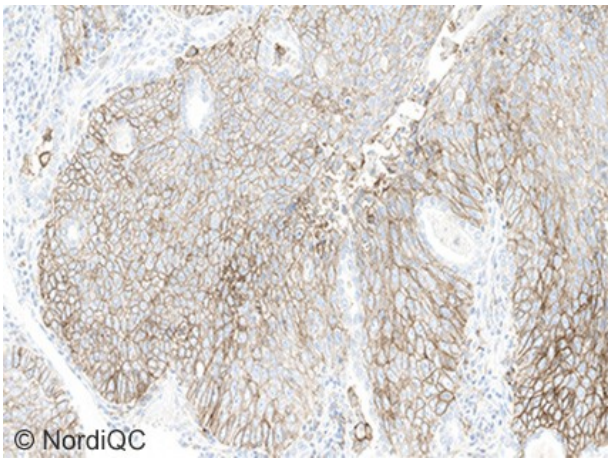


Fig. 5a. Optimal PD-L1 staining result of the NSCLC tissue core no. 19 using same protocol as in Figs. 1a-4a. The majority of neoplastic cells show a moderate to strong membranous staining reaction and classified as TPS high. The patient eligible for 1' line immune therapy.

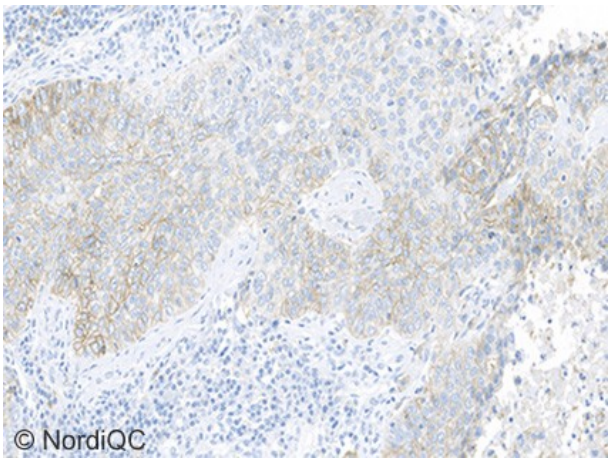


Fig. 5b. Insufficient PD-L1 staining result of the NSCLC tissue core no. 19 (same protocol as in Figs. 1b-4b). Only few cells show a membranous staining reaction.. Few immune cells show a strong staining reaction. The assay has not been calibrated to demonstrate the weak PD-L1 expression in the tumour cells.