Assessment Run C9 2021
PD-L1 TPS/CPS (KEYTRUDA®)

Purpose
This was the ninth assessment for PD-L1 in the NordiQC Companion module. This and the previous assessments for PD-L1 TPS/CPS (KEYTRUDA®) primarily focused on the evaluation of the analytical accuracy of the IHC assays performed by the NordiQC participants to identify patients with NSCLCs and urothelial carcinomas to be treated with KEYTRUDA® as immune therapy. PD-L1 22C3 pharmDx, SK006 and GE006 (Dako/Agilent) and SP263 741-4905 (Ventana/Roche) were used as reference standard methods, and accuracy was evaluated in carcinomas with the dynamic and critical relevant expression levels of PD-L1 characterized and evaluated by TPS and CPS. The obtained score in NordiQC is indicative of the performance of the IHC tests but due to the limited number and composition of samples internal validation and extended quality control, e.g. regularly measuring the PD-L1 results, is needed.

Material
Table 1. **Content of the TMA used for the NordiQC PD-L1 TPS/CPS (KEYTRUDA®) C9 assessment**

	PD-L1 IHC TPS/CPS score*
Tissue controls	
1. Placenta	See section for controls
2-3. Tonsil	See section for controls
Carcinomas	
4. NSCLC	TPS: No; <1%
5. NSCLC	TPS: Low; 15-40%**
6. NSCLC	TPS: Low; 20-40%
7. NSCLC	TPS: High; 90-100%
8. Urothelial carcinoma	CPS: <10
9. Urothelial carcinoma	CPS: ≥10; 80-100 IC***
10. Urothelial carcinoma	CPS: ≥10; 100-200 IC+TC***

* Tumour proportion score (TPS) and combined positive score (CPS) determined by PD-L1 IHC 22C3, SK006, GE006 (Dako/Agilent) and SP263 741-4905 (Ventana/Roche) performed in NordiQC reference lab.
** The tumour showed heterogeneity in the different levels within and in between the TMA's used and focally a TPS ≥50% was observed.
*** IC, Immune cells TC; Tumour cells

All tissues were fixed in 10% neutral buffered formalin.

The participating laboratories were asked to perform the PD-L1 IHC assay for treatment with KEYTRUDA®, evaluate the PD-L1 expression level using the TPS and CPS scoring system and submit the stained slides and scores to NordiQC. This allowed assessment of the technical performance (analytical accuracy) of the PD-L1 TPS/CPS assays and provided information on the reproducibility and concordance of the PD-L1 read-out results among the laboratories.

PD-L1 TPS/CPS, Technical assessment
In order to account for heterogeneity of PD-L1 expression in the individual tumour cores included in the tissue micro array (TMA) blocks, reference slides were made throughout the blocks. The PD-L1 expression levels throughout the blocks were characterized by the CE IVD / FDA approved 22C3 pharmDx kits SK006 (Dako/Agilent) for Autostainer Link 48, CE IVD approved 22C3 pharmDx kit GE006 for Dako Omnis, and also by the CE IVD approved assay (NSCLC, KEYTRUDA®) SP263 741-4905 (Ventana/Roche) for BenchMark in a NordiQC reference laboratory. During the assessment, TPS and CPS categories for each tissue core on the submitted slides were compared to the level in the nearest reference slides.

Criteria for assessing a staining as <u>Optimal</u> include:
The staining is considered perfect or close to perfect in all of the included tissues.
<u>TPS/CPS is concordant to the NordiQC reference data in all carcinomas.</u>

Criteria for assessing a staining as <u>Good</u> include:
The staining is considered acceptable (correct PD-L1 TPS/CPS category) in all of the included tissues. PD-L1 expression in one or more tissues varies significantly from the expected TPS/CPS scores, but still in the correct category. The protocol may be optimized to ensure analytical accuracy.
The technical quality may be improved for e.g. counter staining, morphology and signal-to-noise ratio.
<u>TPS/CPS is still concordant to the NordiQC reference data obtained in all carcinomas</u>.

Criteria for assessing a staining as <u>Borderline</u> include:
The staining is considered insufficient because of a false negative or false positive staining reaction in one of the included carcinomas. The protocol should be optimized.
<u>TPS/CPS is **not** concordant to the NordiQC reference data in one of the carcinomas</u>

Criteria for assessing a staining as <u>Poor</u> include:
The staining is considered very insufficient e.g. because of a false negative or a false positive staining reaction of more than one of the included carcinomas.
Optimization of the protocol is urgently needed.
<u>TPS/CPS is **not** concordant to the NordiQC reference data in two or more of the carcinomas.</u>

An IHC result can also be assessed as **borderline/poor** related to technical artefacts, e.g. poor signal-to-noise ratio, excessive counterstaining, impaired morphology and/or excessive staining compromising the scoring.

Participation

Number of laboratories registered for PD-L1 KEYTRUDA IHC C8	228
Number of laboratories returning PD-L1 KEYTRUDA IHC slides	213 (93%)
Number of laboratories returning PD-L1 scoring sheet	189

Results: 213 laboratories participated in this assessment and returned slides. Two submitted results were excluded due to technical and logistical issues. Of the remaining 211 participants, 82% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Table 2 (see page 3). All slides returned after the assessment were assessed and laboratories received advice if the result was insufficient, but the data were not included in this report.

Performance history
This was the ninth NordiQC assessment of PD-L1 TPS/CPS (KEYTRUDA®). A relatively consistent pass rate has been obtained in the last three assessments, as shown in Graph. 1 below. The number of new participants seems to be consistently increasing about 5% in the last 6 runs.

Graph 1. **Proportion of sufficient results for PD-L1 TPS/CPS (KEYTRUDA®) in the nine NordiQC runs performed.**

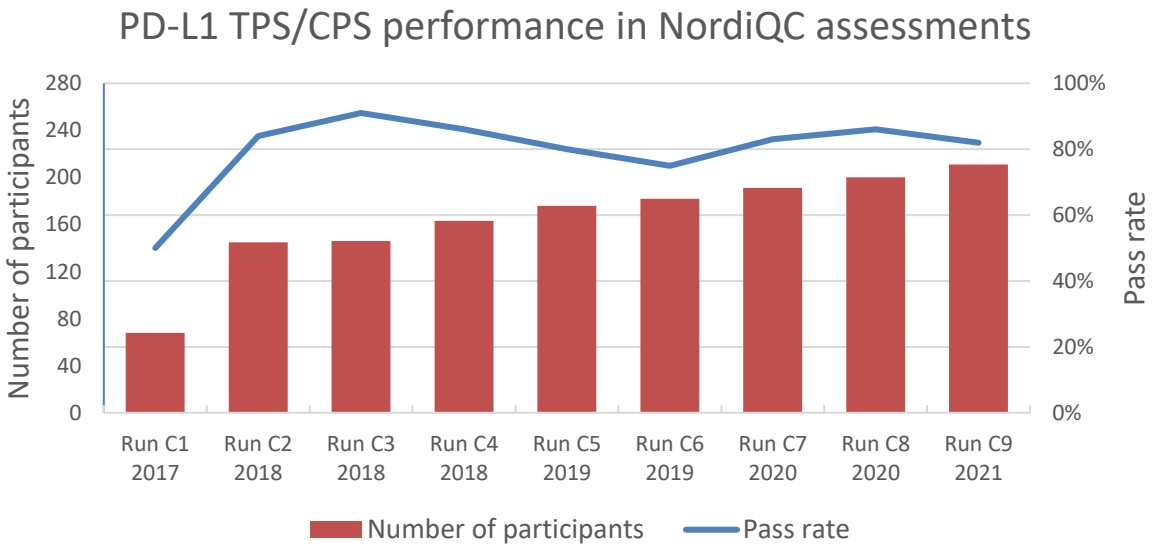


Table 2. **Assessment marks for IHC assays and antibodies run C9, PD-L1 TPS/CPS (KEYTRUDA®)**

CE-IVD / FDA approved PD-L1 assays	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP263, **741-4905 (VRPS)[3]**	42	Ventana/Roche	29	9	4	-	91%	69%
rmAb clone SP263, **741-4905 (LPMS)[4]**	2	Ventana/Roche	-	-	1	1	-	-
rmAb clone SP263, **740-4907 (VRPS)[3]**	13	Ventana/Roche	8	4	1	-	92%	62%
rmAb clone SP142, **740-4859 (VRPS)[3]**	1	Ventana/Roche	-	-	-	1	-	-
mAb clone 22C3 pharmDX, **SK006 (VRPS)[3]**	23	Dako/Agilent	5	14	4	-	83%	22%
mAb clone 22C3 pharmDX, **SK006 (LMPS)[4]**	9	Dako/Agilent	2	3	2	2	56%	22%
mAb clone 22C3 pharmDX, **GE006 (VRPS)[3]**	21	Dako/Agilent	17	4	-	-	100%	81%
mAb clone 22C3 pharmDX, **GE006 (LMPS)[4]**	7	Dako/Agilent	2	3	2	-	71%	29%
rmAb clone 28-8 pharmDX, **SK005 (VRPS)[3]**	2	Dako/Agilent	2	-	-	-	-	-
Antibodies[5] for laboratory developed PD-L1 assays, concentrated antibodies	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
mAb clone **22C3**	38	Dako/Agilent	10	21	5	2	82%	26%
mAb clone **E1L3N**	4	Cell Signaling	-	1	3	-	-	-
rmAb clone **28-8**	1	Abcam	-	1	-	-	-	-
rmAb clone **BSR90**	1	Nordic Biosite	-	1	-	-	-	-
rmAb **CAL10**	3 1	Biocare Zytomed Systems	3	1	-	-	-	-
rmAb clone **QR1**	1	Biocyc	-	-	1	-	-	-
rmAb clone **SP142**	1	Abcam	-	1	-	-	-	-
rmAb clone **ZR3**	1 1	Zeta Corporation Zytomed systems	-	-	2	-	-	-
Ready-To-Use antibodies[6]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP263, **790-4905[6] (VRPS)[3]**	13	Ventana/Roche	8	3	1	1	85%	62%
rmAb clone SP263, **790-4905[6] (LMPS)[4]**	20	Ventana/Roche	12	6	1	1	90%	60%
mAb **405-9A11 PDM572**	1	Diagnostic Biosystems	-	-	1	-	-	-
mAb **IHC441 IHC441-7**	1	GenomeMe	-	-	1	-	-	-
rmAb clone 73-10, **PA0832 (VRPS)[3]**	1	Leica Biosystems	-	1	-	-	-	-
rmAb clone MX070C, **MAB-0854**	2	Maixin	2	-	-	-	-	-
rmAb clone **ZR3 GT228002**	1	Gene Tech	-	-	1	-	-	-
Total	211		100	73	30	8		
Proportion			47%	35%	14%	4%	82%	

1) Proportion of sufficient stains (optimal or good).
2) Proportion of optimal results.
3) Vendor recommended protocol settings – RTU product used in compliance to protocol settings, platform and package insert.
4) Laboratory modified protocol settings for a RTU product applied either on the vendor recommended platform(s) or other platforms.
5) mAb: mouse monoclonal antibody, rmAb: rabbit monoclonal antibody.
6) Ready-To-Use antibodies without predictive claim.

Detailed Analysis
CE IVD / FDA approved assays

SP263 (741-4905, Ventana/Roche): In total, 29 of 44 (66%) protocols were assessed as optimal. This product has a locked protocol on all BenchMark platforms and cannot be changed. The protocol is based on Heat Induced Epitope Retrieval (HIER) in Cell Conditioning 1 (CC1) at 95-100°C for 64 min., 16 min. incubation of primary Ab and OptiView as detection system. Using these protocols settings and applied on BenchMark platform, 38 of 42 (91%) laboratories produced a sufficient staining result (optimal or good).

SP263 (740-4907, Ventana/Roche): In total, 8 of 13 (62%) protocols were assessed as optimal. This product has a locked protocol on BenchMark Ultra platform and cannot be changed. The protocol is based on HIER in CC1 at 95-100°C for 64 min., 16 min. incubation of primary Ab and OptiView as detection system. Using these protocols settings, 12 of 13 (92%) laboratories produced a sufficient staining result.

PD-L1 IHC 22C3 pharmDx (SK006, Dako/Agilent): In total, 7 of 32 (22%) protocols were assessed as optimal. Protocols with optimal results were typically based on the vendor recommended protocol settings based on HIER using EnVision™ FLEX Target Retrieval Solution (TRS) low pH 6.1 at 95-99°C for 20 min. in PT Link, 30 min. incubation of the primary Ab, EnVision FLEX+ as the detection system and performed on Autostainer Link 48. Using these protocol settings, 19 of 23 (83%) laboratories produced a sufficient staining result.
SK006 was frequently used by modified protocol settings e.g. mitigation to other platform as Ventana BenchMark or performed manually with overall inferior performance as shown in Table 2.

PD-L1 IHC 22C3 pharmDx (GE006, Dako/Agilent): In total, 19 of 28 (68%) protocols were assessed as optimal. Protocols with optimal results were typically based on the vendor recommended protocol settings HIER using EnVision™ FLEX TRS low pH 6.1 (GV805) at 95-99°C for 40 min., 40 min. incubation of the primary Ab, EnVision FLEX+ as the detection system and performed on Omnis. Using these protocol settings, 21 of 21 (100%) laboratories produced a sufficient staining result.

Table 3 summarizes the proportion of sufficient and optimal marks for the most commonly used CE IVD / FDA approved assays. 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 3. **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, **741-4905**	38/42 (91%)	29/42 (69%)	-	-
Ventana BenchMark Ultra rmAb SP263, **740-4907**	12/13 (92%)	8/13 (62%)	-	-
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	19/23 (83%)	5/23 (22%)	2/2	1/2
Dako Omnis mAb 22C3 pharmDX, **GE006**	21/21 (100%)	17/21 (81%)	2/2	2/2
Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005**	2/2	2/2	-	-

*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**: 10 of 38 (26%) protocols were assessed as optimal of which six were stained on the Omnis platform (Dako/Agilent), three on the BenchMark Ultra platform (Ventana/Roche) and one on Autostainer (Dako/Agilent).
On Omnis, the protocols providing optimal results were based on a titre of 1:20-30 of the primary Ab, incubation time of 30-40 min., HIER in TRS low pH 6.1 (Dako/Agilent) at 97°C (efficient heating time 30-50 min.) and EnVision FLEX+ as detection system. Using these protocol settings, 8 of 8 (100%) laboratories produced a sufficient staining result.
On BenchMark Ultra, the protocols providing optimal results were based on a titre of 1:40 of the primary Ab, incubation time of 32-64 min., HIER in CC1 (efficient heating time 48-64 min.) and OptiView as detection system. Using these protocol settings, 7 of 7 (100%) laboratories produced a sufficient staining result.

On Autostainer (Dako/Agilent) the protocol providing an optimal result was based on a titre of 1:25 of the primary Ab, incubation time of 40 min., HIER in TRS low pH 6.1 (Dako/Agilent) at 97°C (efficient heating time 20 min.) and EnVision FLEX+ as detection system.

rmAb **CAL10**: 3 of 4 protocols (75%) were assessed as optimal.
One protocol was based on HIER using an alkaline buffer Bond Epitope Retrieval Solution 2 (BERS2, Leica Biosystems) at 99°C for 30 min. The rmAb clone CAL10 (Zytomed) was diluted 1:40, incubated for 20 min. at room temp. and visualized by Leica Refine detection kit and performed on a Leica Biosystems Bond III platform.
One protocol was based on HIER in Diva Decloaker in a pressure cooker at 110°C for 15 min. The rmAb clone CAL10 (Biocare) was diluted 1:200 for 30 min. at room temp. and visualized by MACH4 (Biocare) and performed on a Biocare IntelliPATH platform.
One protocol was based on HIER in Tris/EDTA pH 9 in a pressure cooker at 115°C for 1 min. The rmAb clone CAL10 (Biocare) was diluted 1:50 for 30 min. at room temp. and visualized by EnVision FLEX+ (Dako/Agilent) and performed on a Dako Autostainer Link 48 platform.

Table 4. **Optimal results for PD-L1 for the most commonly used antibody as concentrate on the four main IHC systems***

Concentrated antibodies	**Ventana/Roche BenchMark GX/XT/Ultra**		**Dako/Agilent Autostainer**		**Dako/Agilent Omnis**		**Leica Biosystems Bond III/Max**	
	CC1 pH 8.5	CC2 pH 6.0	TRS pH 9.0	TRS pH 6.1	TRS High pH	TRS Low pH	BERS2 pH 9.0	BERS1 pH 6.0
mAb clone **22C3**	3/9 (67%)	-	-	1/2	0/1	6/8 (75%)	0/4	-

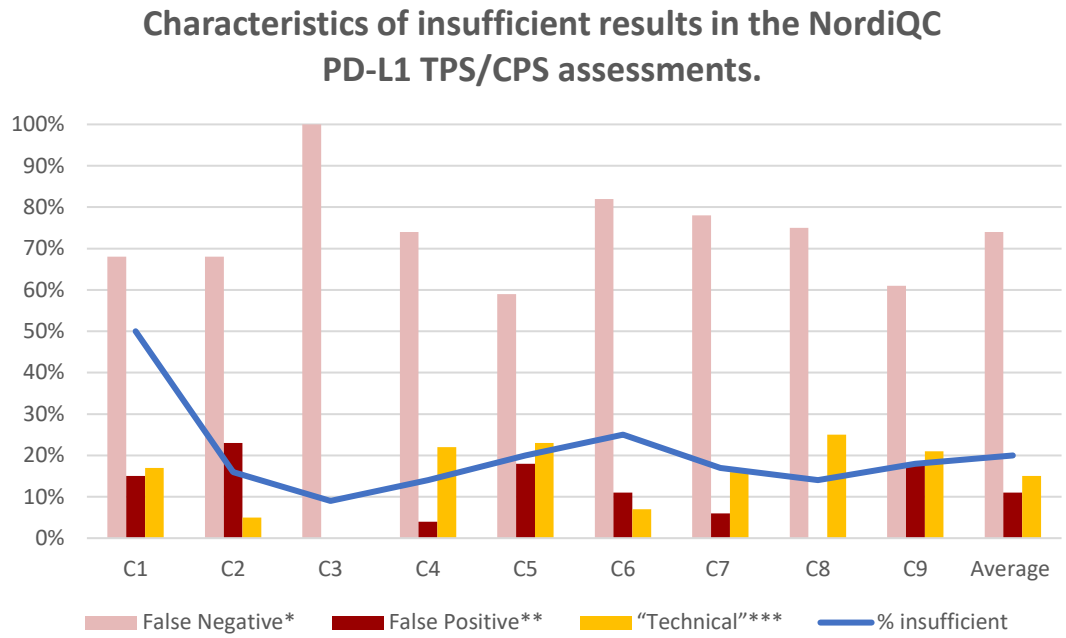
*Antibody concentration applied as listed above, HIER buffers and detection kits used as provided by the vendors of the respective platforms.
**number of optimal results/number of laboratories using this buffer.

Ready-To-Use antibodies for laboratory developed (LD) assays
rmAb **SP263** (790-4905, Ventana/Roche): In total, 20 of 33 (61%) protocols provided an optimal result. Protocols with optimal results were typically based on HIER in CC1 at 95-100°C, efficient heating time 52-64 min., 16 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra or XT. Using these protocols settings, 16 of 19 (84%) laboratories produced a sufficient staining result.

Block construction and assessment reference standards
The tissue micro array (TMA) blocks constructed for this PD-L1 run consisted of 4 NSCLCs, 3 urothelial carcinomas, 2 tonsils and 1 placenta. The NSCLCs were selected to comprise PD-L1 expression levels for each TPS category: TPS negative (<1% PD-L1 positive tumour cells), TPS low (≥1-49%) and TPS high (≥50%). The urothelial carcinomas were selected to comprise 1 carcinoma with CPS<10 and 2 carcinomas with CPS≥10 - one with PD-L1 expression primarily in immune cells and one with PD-L1 expression in both tumour cells and immune cells. Reference slides throughout the individual TMA blocks (interval at each twenty-fifth slide) were stained using the companion diagnostic assays 22C3 pharmDX SK006 (Dako/Agilent), 22C3 pharmDx GE006 (Dako/Agilent) and SP263 741-4905 (Ventana/Roche). 22C3 pharmDX SK006 (Dako/Agilent) was used to characterize PD-L1 for both TPS and CPS levels, whereas 22C3 pharmDx GE006 and SP263 for were mainly used to characterize TPS (reflecting the EU/FDA approved predictive claims for KEYTRUDA® at the assessment). In total, eight identical TMA blocks were constructed and five of these used for this assessment.
Reviewing the reference slides from the blocks, slight heterogenic expression of PD-L1 was seen in one of the tumor cores. In the NSCLC, tissue core no. 5, predominantly scored as TPS low (≥1-49%), focal areas with TPS high ≥50% were identified.
During the assessment, TPS and CPS categories for each tissue core on the submitted slides were compared to the level in the nearest reference slides.
Heterogeneity in PD-L1 expression is well known in NSCLCs and the assessment in this sense emulated clinical settings.

Comments
In this nineth NordiQC assessment for PD-L1 TPS/CPS (KEYTRUDA®), the prevalent feature of an insufficient staining result was a false negative staining result, being observed in 61% (23 of 38) of the insufficient results. As shown in Graph. 2, a false negative staining result has been the most common reason for insufficient staining results in all NordiQC PD-L1 TPS/CPS (KEYTRUDA®) assessments with an average occurrence of 74%. In this run, 21% (8 of 38) of the insufficient results were caused by technical issues as poor-signal-to-noise ratio or excessive cytoplasmic staining reaction compromising the scoring of the PD-L1 status in one or more of the carcinomas. In the remaining 18% (7 of 38) a false positive result was observed.

Graph 2. **Prevalence and characteristics of insufficient results**



* TPS change from high to low or low to negative. And/or CPS change from ≥10 to <10.
** TPS change from negative to low or low to high. And/or CPS change from <10 to ≥10.
*** Interpretation compromised e.g. by poor-signal-to noise ratio, poor morphology, excessive cytoplasmic staining reaction etc.

In this assessment and in concordance with previous runs the majority of insufficient results were related to incorrect TPS categories in one or more of the NSCLCs, whereas the CPS categories of the urothelial carcinomas only were affected in a few cases. PD-L1 IHC demonstration in the NordiQC assessments with combined tumour material has thus been more successful in urothelial carcinomas versus NSCLCs. No plausible reasons for this difference have been identified. The expression levels in the combined tumour materials used for the assessments in combination with different cut-off values and scoring methods might have favoured consistent PD-L1 demonstration in urothelial carcinomas compared to NSCLCs. In order to evaluate IHC accuracy NordiQC strives to include neoplasms with PD-L1 levels close to the critical and clinically relevant thresholds for positivity focusing on both intensity, proportion and subtypes of cells to be scored mimicking real-life diagnostics.

The NSCLC, tissue core no. 5, was most challenging to obtain an optimal result and required a carefully calibrated and reproducible protocol. Virtually all false negative results were seen in the NSCLC, tissue core no. 5, changing the TPS category compared to the level expected and defined by the NordiQC reference standard methods. False positive results were seen in both the NSCLC tissue core no. 5 and/or no. 6, changing the PD-L1 status from TPS low (1-49%) to TPS high (≥50%). Both tumours were characterized to have a TPS in the range of 15-40% as determined by the NordiQC reference standard methods and in addition by the vast majority of participants.

In contrast, virtually all protocols provided the expected PD-L1 status in both the NSCLC, tissue core no. 7, characterized by NordiQC to show a strong membranous staining reaction in all tumour cells and the urothelial carcinoma, tissue core no. 10 with CPS≥10 expressed in both immune cells and tumour cells. The NSCLC, tissue core no. 4, and the urothelial carcinoma, tissue core no. 8, were consistently negative by all protocols submitted.

Similar to run C8 and compared to previous runs an increasing proportion (35%) of the participants obtained a score as "Good". In 60% of these (44 of 73), this was due to a general weak staining result or a reduced TPS/CPS, but with no change of TPS/CPS-category in any of the carcinomas and thus still an accurate PD-L1 status for treatment decision. In 7% (5 of 73) an increased TPS/CPS score was observed compared to the level expected, but again without any change in the PD-L1 status. 16% (12 of 73) of the results assessed as "Good" were characterized by poor signal-to-noise ratio, impaired morphology, too weak or excessive counterstaining.
And finally, in 4% (3 of 73) a coarse granular staining reaction compromising the evaluation of the membranous staining reaction was observed. This pattern was only seen for protocols based on OptiView with amplification kit and the general use of these protocol settings being reduced compared to previous assessments.

The Ventana/Roche PD-L1 IHC assays 741-4905 and 740-4907 for BenchMark (Ultra/XT/GX) with predictive claims, based on the SP263 clone, were used by 27% of the participants and in total provided a pass rate of 91%, 67% optimal when applied by protocol settings in compliance with vendor recommendations (see Table 3). The assays are locked for central protocol settings and based on HIER in CC1 for 64 min., incubation in primary Ab for 16 min. and use of OptiView as detection system. Despite the locked protocol conditions for the two assays, some laboratories submitted protocols with reported modified settings typically indicating a change for HIER and/or incubation time of primary Ab. The different protocol settings submitted were disregarded for the two assays product no. 741-4905 and 740-4907 in this report and all protocols thus compiled as used by vendor recommended protocol settings as shown in Tables 2 and 3.

The Dako/Agilent 22C3 pharmDx assay SK006 for Autostainer Link 48 was used by 15% of the participants and provided a pass rate of 83% and 22% optimal when applied by protocol settings in compliance with vendor recommendations (see Table 3).
The pass rate and in particular the proportion of optimal results was reduced compared to the levels seen in previous NordiQC assessments and also to both the corresponding Dako/Agilent 22C3 pharmDx assay GE006 for Omnis and the Ventana/Roche PD-L1 CDx assays based on SP263. The results assessed as "Good" were all characterized either by a generally weak staining reaction and/or a reduced number of cells being demonstrated compared to the level expected, but still no impact on PD-L1 status in any of the carcinomas. The insufficient results were caused by a significantly reduced proportion of cells demonstrated and a change of the PD-L1 status in one or more of the carcinomas. No plausible reason for the reduced analytical sensitivity and accuracy could be identified.

The Dako/Agilent 22C3 pharmDx assay GE006 for Omnis was used by 13% of the participants and provided a pass rate of 100% and 81% optimal when applied by protocol settings in compliance with vendor recommendations (see Table 3).
Similar to the data generated in runs C6, C7 and C8 it was observed that the PD-L1 22C3 GE006 assay for Omnis was more successful compared to 22C3 pharmDx SK006. Cumulated data for the 4 successive runs has shown a pass rate of 100% (61 of 61) for laboratories using GE006 by vendor recommended protocol settings. In comparison a pass rate of 80% (60 of 75) for laboratories using SK006 has been obtained. The different pass rates observed have to be taken with caution due to relatively few data observations, but a clear trend so far has been observed in the four successive runs performed.
In this context it has to be emphasized that the 22C3 GE006 assay for Omnis is only validated by the vendor for PD-L1 status and predictive claim in NSCLC with TPS as scoring system and at present not validated for any indication with CPS as scoring system including urothelial carcinoma.

The Dako/Agilent pharmDx SK005 28-8 for Autostainer Link 48 was used by two laboratories. Both used the recommended protocol settings and both results being assessed as optimal.

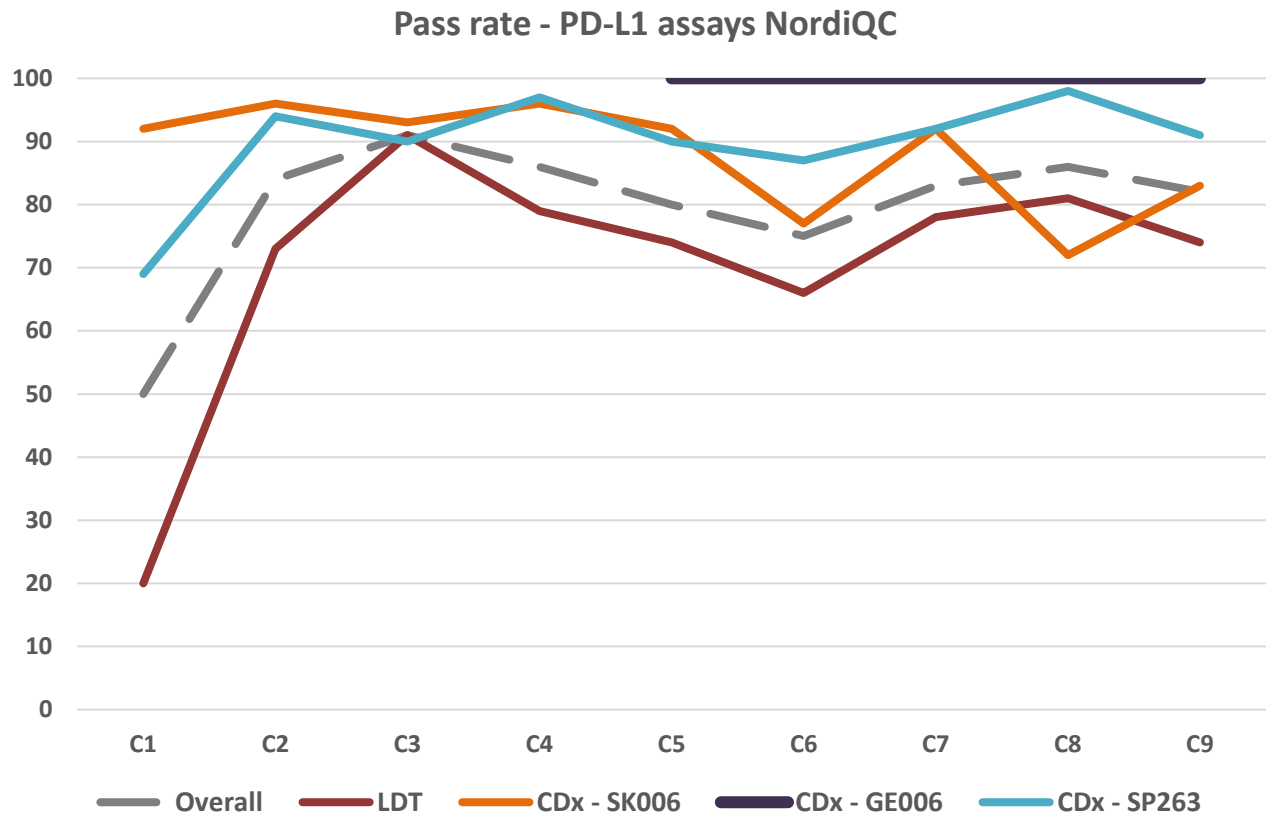
Grouped together, and using vendor recommended protocol settings, the five above mentioned CE IVD approved PD-L1 IHC assays with predictive claims - irrespective of indication and drug associated - provided a pass rate of 91% (92 of 101) and 60% being optimal (61 of 101). These levels indicate a possibility of interchangeability between the assays for PD-L1 status for KEYTRUDA® using the present cut-off values and scoring methods for TPS/CPS in the two indications addressed in this module. This must be validated by end-user according to local regulations.

The Ventana CDx assay based on SP142 was used by one participant and the result submitted assessed as insufficient being concordant to observations and data for the SP142 CDx assay in previous NordiQC runs for PD-L1 TPS/CPS. Several publications inclusive Blueprint studies 1 and 2 (Hirsch, Tsao et al) have indicated poor analytical concordance for SP142 compared to the other CDx assays for TPS.

Laboratory developed (LD) assays either based on a concentrated Ab, an RTU Ab without any predictive claim or a companion diagnostic assay not used strictly accordingly to vendor recommendations were applied by 52% (109 of 211) of the participants. For this group a pass rate of 74% was observed, 26% being optimal and the performance was overall inferior to the CDx assays being used as "plug-and-play".

The performance of most commonly used IHC CDx and LD assays for PD-L1 is summarized and shown in Graph 3 below.

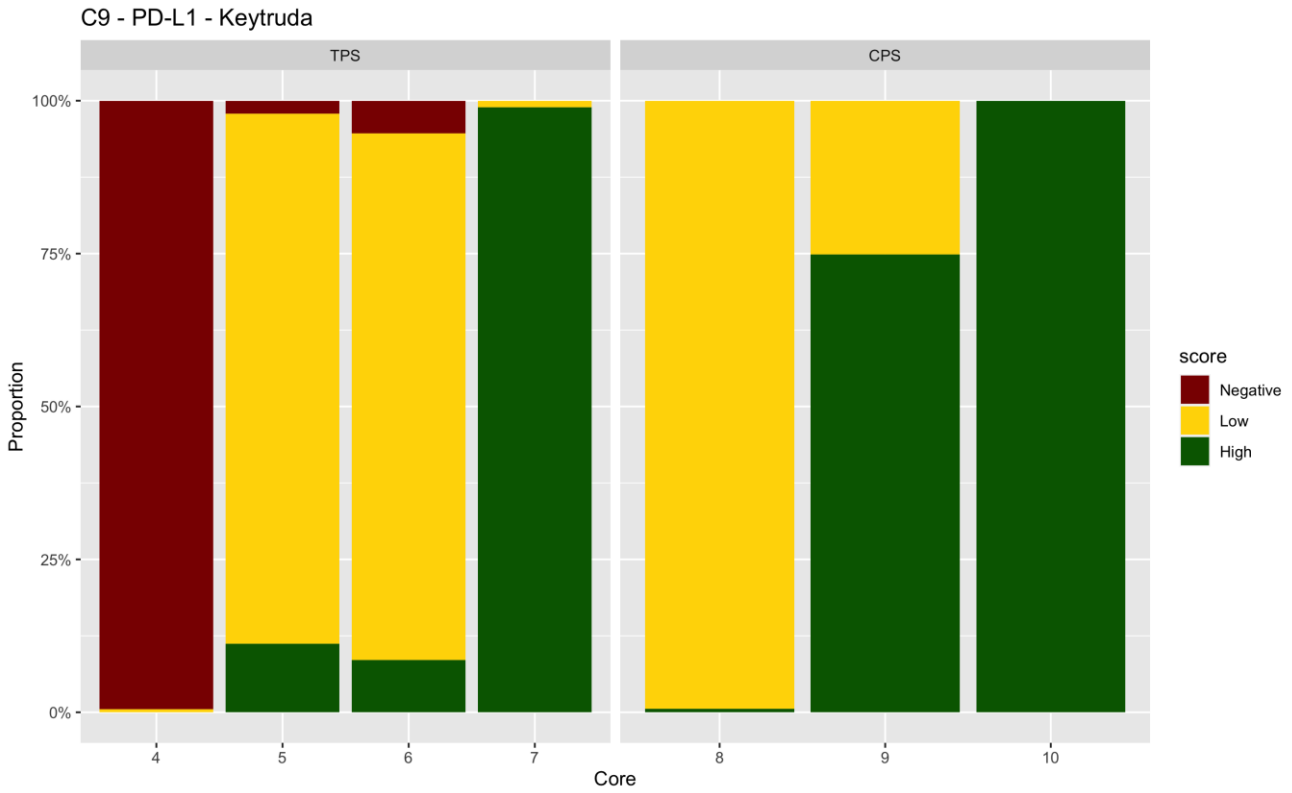
Graph 3. **Proportion of pass rates for PD-L1 TPS/CPS assays in the nine NordiQC runs performed**



The mAb clone 22C3 was the most widely used concentrated Ab within a LD assay (n=39) providing a pass rate of 82%, 26% optimal.
As described above for optimal protocol settings for mAb clone 22C3 as concentrated format, successful and interlaboratory reproducible settings have been identified for BenchMark (Ventana/Roche) and Omnis (Dako/Agilent), whereas the performance on BOND III / BOND MAX (Leica Biosystems) has shown to be inferior. Cumulated data for run C8 and C9 focusing on the performance of mAb clone 22C3 on the BOND platforms have shown a pass rate of 22% (2 of 9), no optimal, despite the clone 22C3 was applied by similar central protocol settings on BOND compared to both BenchMark and Omnis, but so far with limited success. Only few data observations generated and conclusions to be taken with caution.

As indicated in run C8, RTU Abs without predictive claims are now categorized separately in Table 2 and includes the Ventana rmAb clone SP263, 790-4905 previously being listed as CE IVD /FDA approved PD-L1 assay. The RTU format is similar to the two other Ventana SP263 products but positioned as an open "analytical product" for PD-L1 and hereby option to validate the protocol by the end-user.
Laboratories using the RTU format either by protocol settings in compliance with the vendor recommendations or by modified settings obtained a pass rate and proportion of optimal results very similar to the levels seen for the "locked PD-L1 assays" 741-4905 and 740-4907 (see Table 1).

PD-L1 interpretation and scoring consensus:
Participants were asked to score each of the cores using either tumour proportion score (TPS) for the NSCLCs or combined positive score (CPS) for the urothelial carcinomas.



Graph 1. **NordiQC PD-L1 run C9: Tumour Proportion scores (TPS) in NSCLCs (core no. 4-7) and Combined Positive Score (CPS) in urothelial carcinomas (core no. 8-10).**

As seen in Graph 1, a relatively high consensus rates were observed for the tissue core 4-8 and 10, whereas the consensus rates were significantly lower in tissue core 9.

When stratifying for assessment marks, analysis indicated that participants that had received an insufficient mark (borderline or poor) for the technical assessment of their PD-L1 result also had a higher tendency to perform an incorrect read-out of TPS and/or CPS in the submitted slides.

Controls
Tonsil and placenta were used as positive and negative tissue controls. In this and previous assessments, tonsil was found to be superior to placenta, as tonsil displayed a dynamic and clinical relevant range of PD-L1 expression levels, whereas placenta virtually only contained cells (trophoblasts) with high level PD-L1 expression.
In tonsil, protocols with optimal results typically provided the following reaction pattern:
A moderate to strong predominantly membranous staining reaction in dispersed crypt epithelial cells, a weak to moderate, typically punctuated membranous staining reaction of the majority of germinal centre macrophages and scattered interfollicular lymphocytes and macrophages. No staining reaction in the vast majority of lymphocytes and normal stratified squamous epithelial cells.
It was observed that rmAb SP263 (741-4905, 790-4905/4907, Ventana/Roche) typically provided a higher proportion of positive inter- and intra-follicular immune cells compared to the Dako/Agilent 22C3 PD-L1 assays (SK006 and GE006). For other clones, e.g. mAb clone CAL10 typically a stronger staining reaction in more germinal centre macrophages were seen compared to mAb clone 22C3, when the clones provided otherwise optimal and accurate results in the carcinomas. This emphasizes that the expected test performance characteristics in tonsil must be correlated to the PD-L1 IHC test/clone used both for the inter- and intra-PD-L1 IHC reproducibility evaluation.

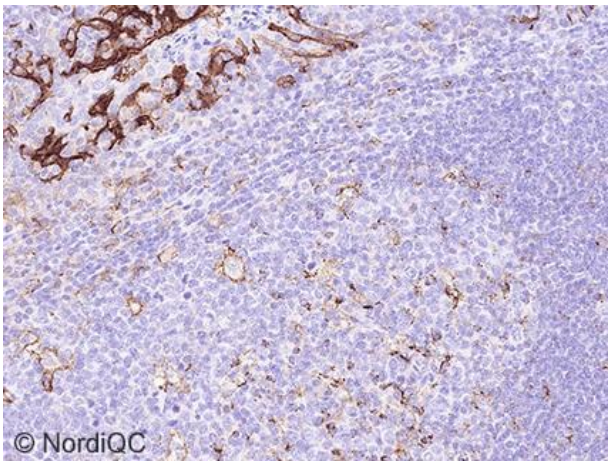


Fig. 1a
Optimal staining result of tonsil using the PD-L1 IHC 22C3 pharmDx kit, GE006, Dako/Agilent on Omnis following the vendor recommended protocol settings.
A weak to moderate, but distinct punctuated membranous staining reaction of germinal centre macrophages and dispersed lymphocytes is seen.
Crypt epithelial cells show a strong staining reaction.
No staining reaction is seen in the vast majority of lymphocytes.
Also compare with Figs. 2a – 5a, same protocol.

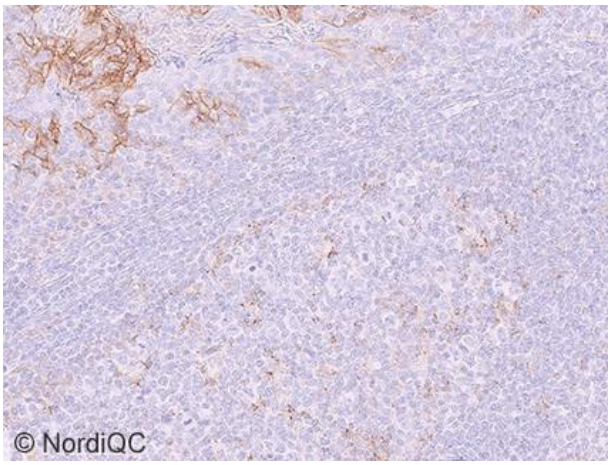


Fig. 1b
Insufficient staining result of tonsil, using the mAb clone 22C3 within a laboratory developed test for PD-L1.
The epithelial crypt cells with high level PD-L1 expression are distinctively demonstrated, whereas the proportion, intensity and distinct membranous staining reaction of germinal centre macrophages and lymphocytes being decreased indicating a reduced analytical sensitivity of the IHC protocol applied – same field as Fig. 1a.
The protocol provided an overall insufficient result characterized by a too low level of analytical sensitivity for PD-L1, as shown in Figs. 3b and 4b, same protocol.

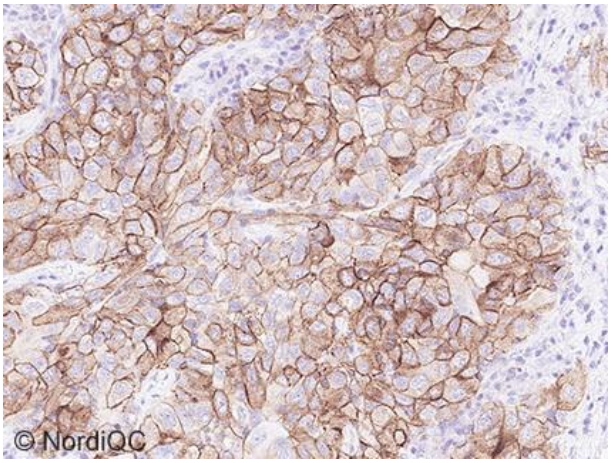


Fig. 2a
Optimal staining result of the NSCLC, tissue core no. 7, using the same protocol as in Fig. 1a.
Virtually all tumour cells show a moderate to strong membranous staining reaction.
The tumour was categorized as TPS high (≥50%) and thus eligible for first line immune therapy with KEYTRUDA® (different regional cut-offs occur).

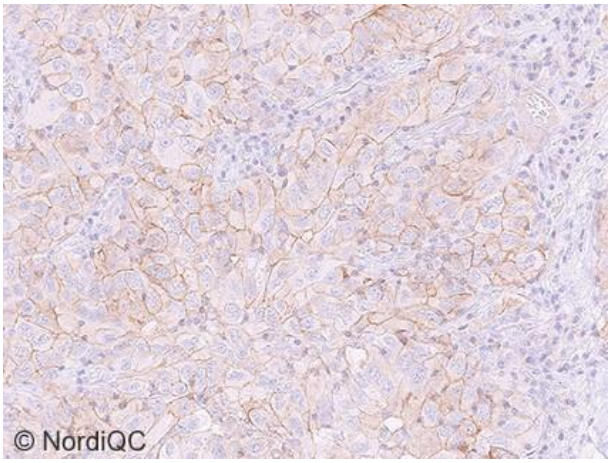


Fig. 2b
Staining result of the NSCLC, tissue core no. 7, using the same protocol as in Fig. 1b
The vast majority of tumour cells show a weak to moderate membranous staining reaction.
The tumour was despite a reduced intensity of the tumour cells still categorized as TPS high (≥50%) and thus eligible for first line immune therapy with KEYTRUDA® (different regional cut-offs occur).
However, also compare with Figs. 3b and 4b, same protocol.

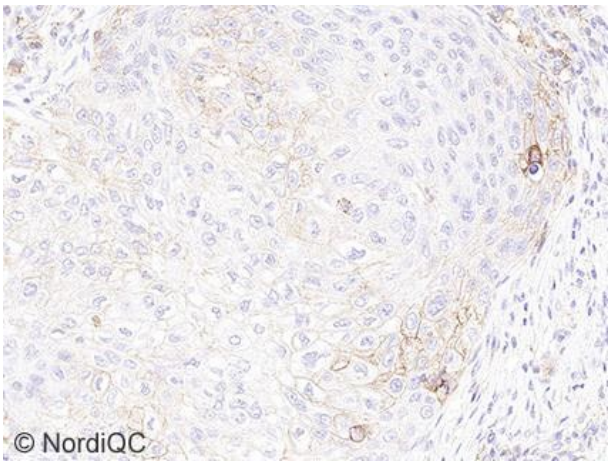


Fig. 3a
Optimal staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1a and 2a.
Approximately 20% of the tumour cells show a weak to moderate membranous staining reaction.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA® (different regional cut-offs occur).

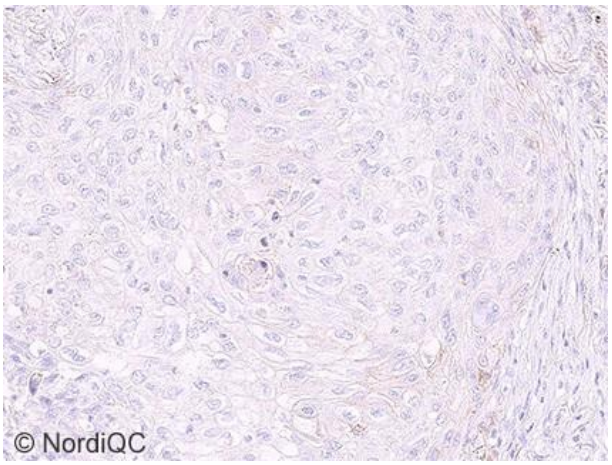


Fig. 3b
Insufficient staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1b and 2b.
<1% of tumour cells show a membranous staining reaction changing the TPS category from the expected low to negative – same field as Fig. 3a

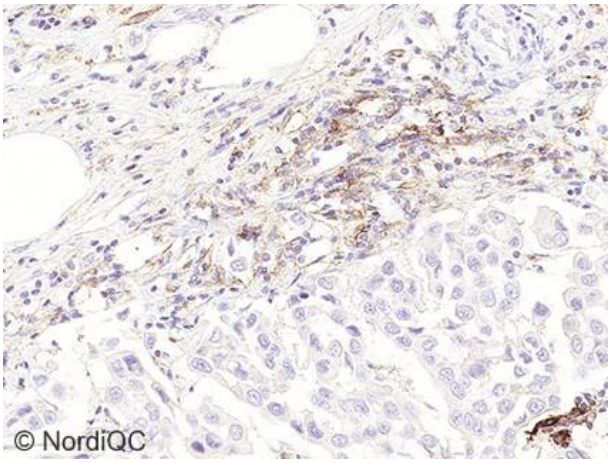


Fig. 4a
Optimal staining result of the urothelial carcinoma, tissue core no. 9, using the same protocol as in Figs. 1a - 3a.
A weak to moderate staining reaction is seen in lymphocytes and macrophages adjacent to the tumour cells being negative.
The tumour was categorized as CPS ≥10 and thus eligible for immune therapy with KEYTRUDA®

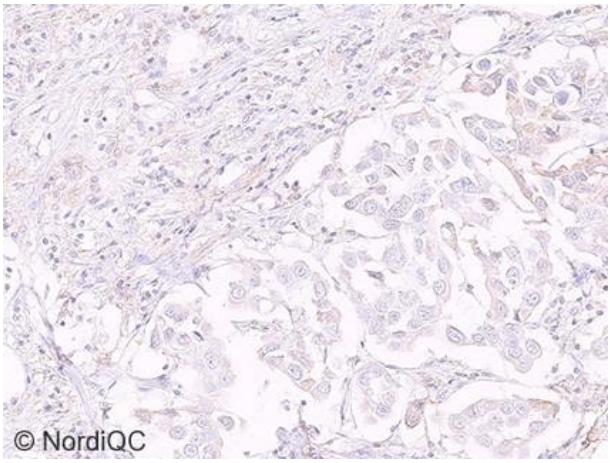


Fig. 4b
Insufficient staining result of the urothelial carcinoma, tissue core no. 9, using the same protocol as in Figs. 1b – 3b.
Virtually no staining reaction in immune cells or tumour cells is seen and the PD-L1 category changed from the expected CPS ≥10 to CPS <10 and not being eligible for immune therapy.

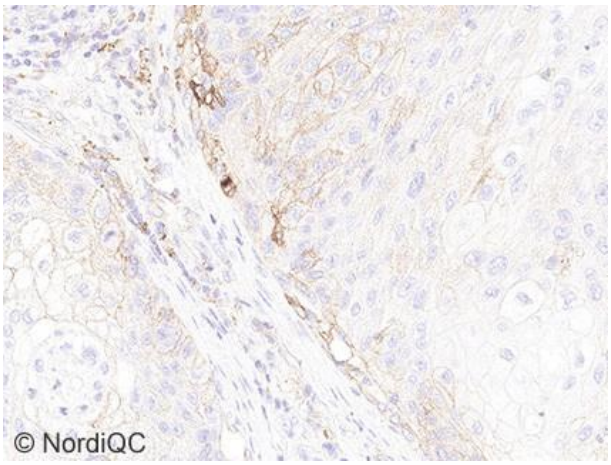


Fig. 5a
Optimal staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1a - 4a.
Approximately 20% of the tumour cells show a weak to moderate membranous staining reaction.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA® (different regional cut-offs occur).

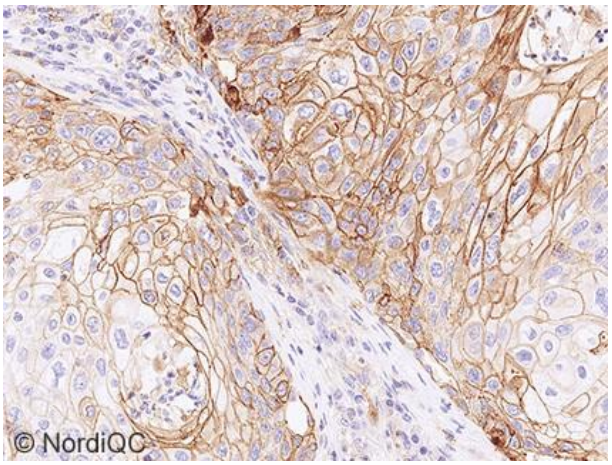


Fig. 5b
Insufficient staining result of the NSCLC, tissue core no. 5, using the rmAb ZR3 by inappropriate protocol settings providing a too high level of the analytical sensitivity for PD-L1.
This tumour was expected to be TPS low as defined by the NordiQC reference standard methods and virtually all participants.
By the applied protocol virtually all tumour cells are positive changing the TPS category from low to high.

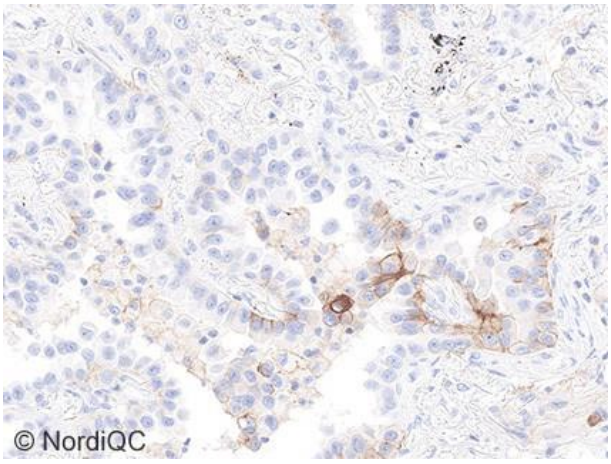


Fig. 6a
Optimal staining result in the NSCLC, tissue core no. 6, using the CDx assay SP263, 741-4905 Ventana/Roche on BenchMark Ultra following the vendor recommended protocol settings.
Approximately 20% of the tumour cells show a weak to moderate membranous staining reaction.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA® (different regional cut-offs occur).

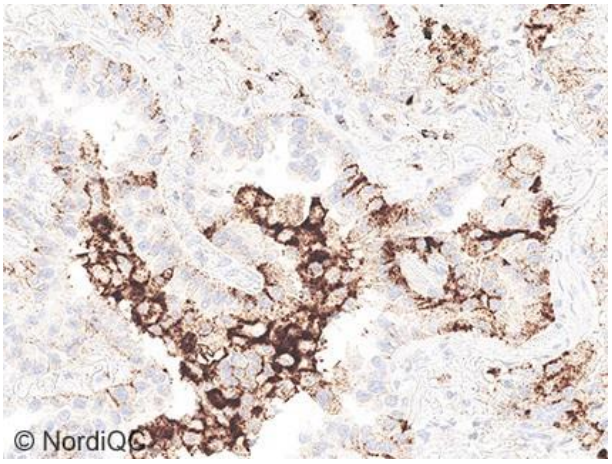


Fig. 6b
Insufficient staining result in the NSCLC, tissue core no. 6, using the CDx assay 22C3, GE006 Dako/Agilent off-label on BenchMark Ultra by laboratory modified protocol settings based on OptiView + Amplification kit as detection system.
The staining pattern of PD-L1 is observed as a coarse granular precipitation and difficult to differentiate if being seen in immune cells or tumour cells. In addition the number of tumour cells demonstrated is increased and PD-L1 category changed from TPS low to high – same field as in Fig. 6a.

SN/LE/HLK 06.07.2021