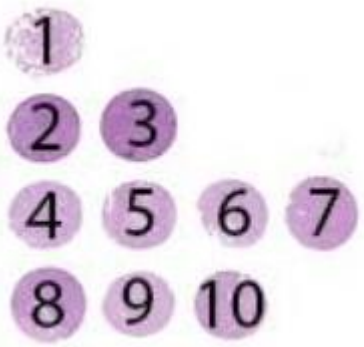
Assessment Run C11 2022
PD-L1 TPS/CPS

Purpose
This was the eleventh assessment for PD-L1 in the NordiQC Companion module. This assessment 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 non-small cell lung cancer (NSCLC) and triple negative breast carcinoma (TNBC) to be treated with KEYTRUDA® as immune therapy. PD-L1 22C3 pharmDx, (Dako/Agilent) was used as reference standard method, and accuracy was evaluated in carcinomas with the dynamic and critical relevant expression levels of PD-L1 characterized 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 additional internal validation/verification and extended quality control, e.g. regularly measuring the PD-L1 results, is needed.
This was the first assessment for PD-L1 TPS/CPS comprising TNBCs being integrated in the material circulated on the expense of urothelial carcinomas (same cut-off's and scoring methods for the two entities.

Material
Table 1. **Content of the TMA used for the NordiQC PD-L1 TPS/CPS (KEYTRUDA®) C11 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; 5-30%***
6. NSCLC	TPS: High; 90-95%
7. NSCLC	TPS: High; 100%
8. TNBC	CPS: <10****
9. TNBC	CPS: ≥10; 30-40 IC#
10. TNBC	CPS: ≥10; 100 TC# + IC#



* Tumour proportion score (TPS) and combined positive score (CPS) determined by PD-L1 IHC 22C3, pharmDx (Dako/Agilent) performed in NordiQC reference lab.
** Focally in one of the six TMA's used for the assessment, areas with TPS 1-3% were observed
*** The tumour showed heterogeneity in the different levels within and in between the TMA's used.
**** Focally in one of the six TMA's used for the assessment, areas with CPS≥10 were observed
# IC, Immune cells - TC; Tumour cells

All tissues were fixed in 10% neutral buffered formalin.

The participating laboratories were asked to perform their PD-L1 IHC assay for treatment decision 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 were thus characterized in every twenty-fifth slide and during the assessment, TPS and CPS categories for each tissue core on the submitted slides from the participants were compared to the level in the nearest reference slide.

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.

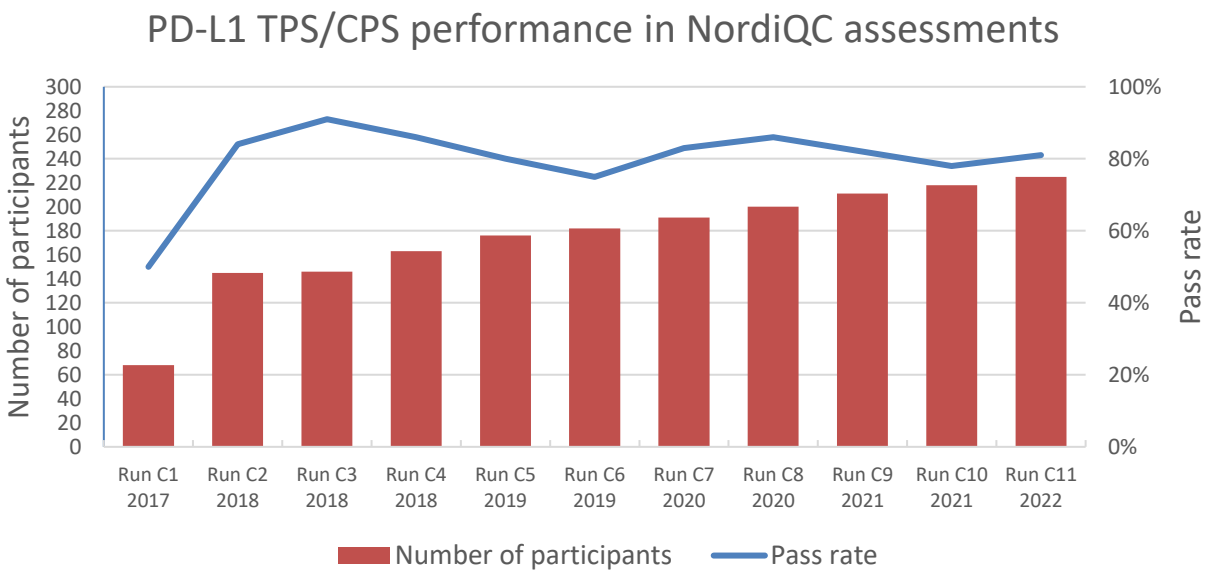
Participation

Number of laboratories registered for PD-L1 KEYTRUDA IHC C11	246
Number of laboratories returning PD-L1 KEYTRUDA IHC slides	225 (92%)
Number of laboratories returning PD-L1 scoring sheet	207

Results
225 laboratories participated in this assessment and returned slides. 81% of the participants 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 eleventh NordiQC assessment of PD-L1 TPS/CPS (KEYTRUDA®). A relatively consistent pass rate has been observed in the latest runs as shown in Graph. 1 below. The number of new participants seems to be consistently increasing about 3-5% in each run.

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



Conclusion

This was the eleventh NordiQC assessment of PD-L1 for TPS/CPS status with focus on NSCLCs and TNBCs. 225 laboratories participated and a pass rate of 81% was observed.

The PD-L1 IHC pharmDx assay, 22C3 GE006, Dako/Agilent applied in concordance to the vendor recommended guidelines, was most successful providing a pass rate of 97%, 81% optimal, being superior to both other companion diagnostic assays and LD assays based on concentrated Abs or RTU systems without predictive claim. LD assays based on mAb clone 22C3 and performed on the fully automated platforms BenchMark and Omnis provided highly reproducible results with "an all time high" pass rate of 95%.

In this run and similar to observations seen in previous NordiQC runs PD-L1 TPS/CPS, the insufficient PD-L1 IHC results were most frequently characterized by a reduced proportion of PD-L1 positive cells compared to the level expected and defined by the NordiQC reference standard methods.

Table 2. **Assessment marks for IHC assays and antibodies run C11, 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]**	50	Ventana/Roche	9	24	16	1	66%	18%
rmAb clone SP263, **741-4905 (LPMS)4**	1	Ventana/Roche	-	-	-	1	-	-
rmAb clone SP263, **740-4907 (VRPS)[3]**	11	Ventana/Roche	3	5	1	2	73%	27%
rmAb clone SP142, **741-4860 (VRPS)[3]**	1	Ventana/Roche	-	-	-	1	-	-
mAb clone 22C3 pharmDX, **SK006 (VRPS)[3]**	24	Dako/Agilent	14	6	4	-	83%	58%
mAb clone 22C3 pharmDX, **SK006 (LMPS)[4]**	13	Dako/Agilent	4	5	3	1	69%	31%
mAb clone 22C3 pharmDX, **GE006 (VRPS)[3]**	31	Dako/Agilent	25	5	1	-	97%	81%
mAb clone 22C3 pharmDX, **GE006 (LMPS)[4]**	9	Dako/Agilent	4	5	-	-	100%	44%
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**	39	Dako/Agilent	19	18	2	-	95%	49%
rmAb **CAL10**	4 2	Zytomed Systems Biocare Medical	1	3	2	-	67%	17%
rmAb clone **E1L3N**	5	Cell Signaling	3	-	2	-	60%	60%
rmAb clone **QR1**	1 1	Quartett Diagomics	-	2	-	-	-	-
Ready-To-Use antibodies[6]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
mAb clone C9C9 **CPM-0278**	1	Celnovte	-	1	-	-	-	-
rmAb clone SP263, **790-4905[6] (VRPS)[3]**	17	Ventana/Roche	3	13	1	-	94%	18%
rmAb clone SP263, **790-4905[6] (LMPS)[4]**	9	Ventana/Roche	-	7	2	-	78%	-
rmAb clone AC37 **AD80167**	1	Abcarta	-	-	1	-	-	-
rmAb clone QR1 **2-PR292-13**	2	Diagomics	-	1	-	1	-	-
rmAb clone RM320 **8263-C010**	1	Sakura Finetek	1	-	-	-	-	-
Total	225		85	98	35	7		
Proportion			38%	43%	16%	3%	81%	

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, 9 of 50 (18%) 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 100°C for 64 min., 16 min. incubation of primary Ab and OptiView as detection system. Using these protocols settings and applied on the BenchMark platform, 30 of 46 (65%) laboratories produced a sufficient staining result (optimal or good).

SP263 (740-4907, Ventana/Roche): In total, 3 of 11 (27%) 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 100°C for 64 min., 16 min. incubation of primary Ab and OptiView as detection system. Using these protocols settings, 8 of 11 (73%) laboratories produced a sufficient staining result.

PD-L1 IHC 22C3 pharmDx (SK006, Dako/Agilent): In total, 18 of 37 (49%) 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, 20 of 24 (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 an overall inferior performance as shown in Table 2.

PD-L1 IHC 22C3 pharmDx (GE006, Dako/Agilent): In total, 29 of 40 (73%) 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, 30 of 31 (97%) 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**	33/50 (66%)	9/50 (16%)	-	-
Ventana BenchMark Ultra rmAb SP263, **740-4907**	8/11 (73%)	3/11 (27%)	-	-
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	20/24 (83%)	14/24 (58%)	1/2	1/2
Dako Omnis mAb 22C3 pharmDX, **GE006**	30/31 (97%)	25/31 (81%)	4/4	3/4
Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005**	2/2	02	-	-

*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**: In total, 19 of 39 (49%) protocols were assessed as optimal of which 14 were stained on the BenchMark Ultra platform (Ventana/Roche), three on the Omnis platform (Dako/Agilent), one on Autostainer Link 48 (Dako/Agilent) and one manually.
On BenchMark Ultra, the protocols providing optimal results were based on a titre of 1:20-50 of the primary Ab, incubation time of 32-120 min., HIER in CC1 (efficient heating time 32-64 min.) and OptiView as detection system. Using these protocol settings, 17 of 18 (94%) laboratories produced a sufficient staining result.
On Omnis, the protocols providing optimal results were based on a titre of 1:20-40 of the primary Ab, incubation time of 30-40 min., HIER in TRS low pH 6.1 at 97°C (efficient heating time 30-50 min.) and EnVision FLEX+ as detection system. Using these protocol settings, 7 of 7 (100%) laboratories produced a sufficient staining result.

rmAb **E1L3N**: 3 of 5 protocols (60%) were assessed as optimal.
The three protocols were based on HIER using an alkaline buffer Bond Epitope Retrieval Solution 2 (BERS2, Leica Biosystems) at 99°C for 30 min. The mAb clone E1L3N was diluted 1:100-500, incubated for 30-60 min. at room temp. and visualized by Leica Refine detection kit and performed on a Leica Biosystems Bond III platform. Using these or similar protocol settings, 3 of 4 (75%) laboratories produced a sufficient staining result.

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**	14/18** (78%)	-	-	1/4	-	3/5 (60%)	0/2	-

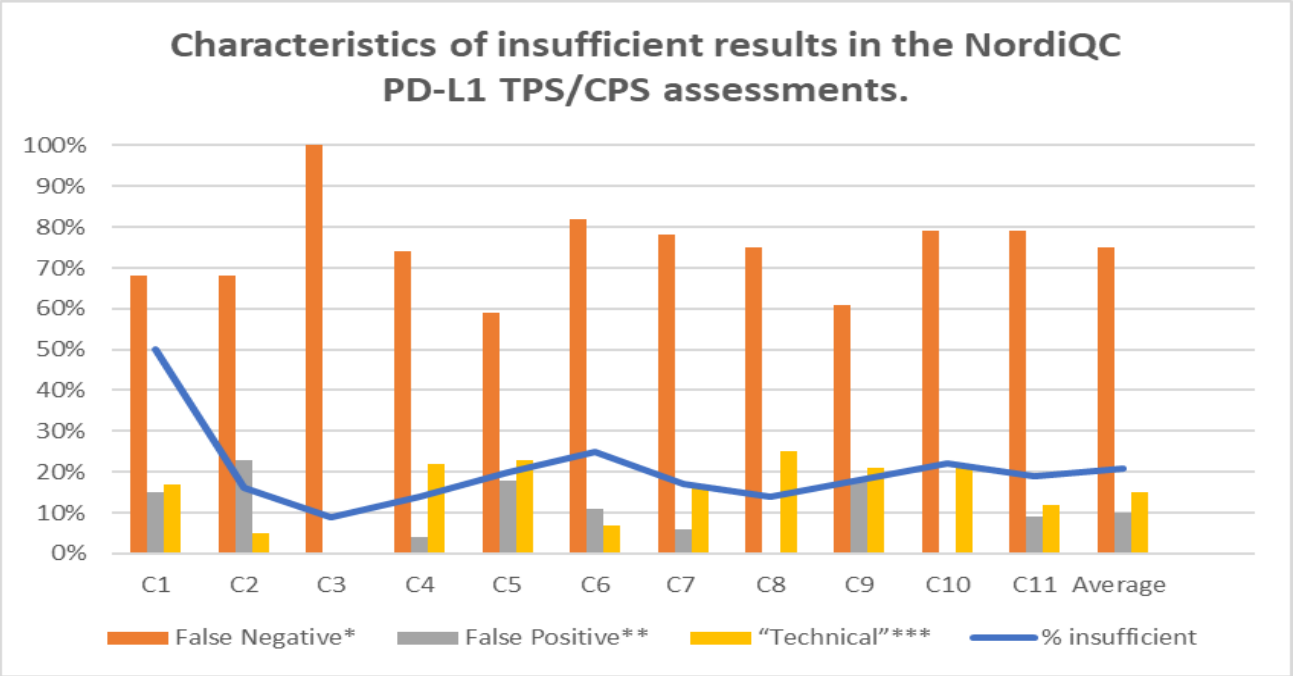
*Antibody concentration applied as listed above, HIER buffers and detection kits used as provided by the vendors of the respective platforms.
**number of optimal results/number of laboratories using this buffer.

Ready-To-Use antibodies for laboratory developed (LD) assays
rmAb **SP263** (790-4905, Ventana/Roche): In total, 3 of 26 (12%) 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., 16-20 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra or GX. Using these protocols settings, 17 of 19 (89%) 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 four NSCLCs, three TNBCs, two tonsils and one 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 TNBCs were selected to comprise one carcinoma with CPS<10 and two 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 assay 22C3 pharmDX (Dako/Agilent).
In total, eight identical TMA blocks were constructed and six of these used for this assessment.
Reviewing the reference slides from the blocks, a 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 with a positivity level in the range of 5-30%, focal areas with an increased level close to TPS high ≥50% and also areas with TPS neg <1% 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 eleventh 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 79% 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 75%. 9% of the insufficient results were related to a false positive staining result while the remaining 12% of the insufficient results were caused by technical issues as poor-signal-to-noise ratio, excessive cytoplasmic staining reaction or a coarse and indistinct granular staining reaction compromising the scoring of the PD-L1 status in one or more of the carcinomas.

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



* TPS changes from high to low or low to negative. And/or CPS changes from ≥10 to <10.
** TPS changes from negative to low or low to high. And/or CPS changes 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 TNBCs only were affected in a few cases. This observation was fully identical to the results obtained and described in previous NordiQC PD-L1 TPS/CPS assessments with the combination of NSCLCs and urothelial carcinomas. Overall, it has thus been observed that the PD-L1 IHC demonstration in the NordiQC assessments with combined tumour material has thus been more successful and accurate in non-NSCLC carcinomas as TNBCs and 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 the non-NSCLCs 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 two NSCLCs, tissue cores no. 5 and 6, characterized as TPS low and TPS high by the NordiQC reference standard methods, respectively, were most challenging to obtain an optimal result. Virtually all false negative results were as such seen in one or both of these NSCLCs, changing the TPS category compared to the level expected and defined by the CE IVD approved PD-L1 IHC assays used as the NordiQC reference standard methods. In addition, the TNBC, tissue core no. 9, expected to show a CPS≥10 (range 30-40) with a positive staining reaction primarily in the immune cells, also was found to be challenging and typically also was false negative when one or both of the two NSCLCs showed an inferior result.

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 TNBC, tissue core no. 10, with CPS≥10 expressed in both immune cells and tumour cells. The false positive cases were only observed in the NSCLCs and seen in tissue core no. 4 and 5 changing the TPS status from negative to low and low to high, respectively.

43% (n=98) of the results submitted were marked as "Good". In 78% of these (76 of 98), this was due to a significantly reduced TPS/CPS level, but with no change of the TPS/CPS-category in any of the carcinomas and thus still an accurate PD-L1 status for treatment decision. In 3% (3 of 98) an increased TPS/CPS level was observed compared to the level expected, but again without any change in the TPS/CPS-category and PD-L1 status. In the remaining 19% (19 of 98) of the results assessed as "Good" these were characterized by poor signal-to-noise ratio, impaired morphology, too weak or excessive counterstaining and/or a coarse granular staining reaction compromising the evaluation of the membranous staining reaction. The latter only seen for protocols based on OptiView with amplification kit (Ventana/Roche).

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 28% of the participants and in total provided an overall pass rate of 67%, 20% 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.
In concordance to the latest PD-L1 TPS/CPS assessment run C10 but in contrast to previous assessments, the two Ventana/Roche PD-L1 assays based on the rmAb clone SP263 provided an inferior performance caused by a reduced analytical sensitivity in one or more of the tissues included. As described above the reduced analytical sensitivity was primarily seen in the NSCLCs tissue core no. 5 and the TNBC tissue core no. 9 changing the PD-L1 category. However also the TNBC tissue core no. 10 was affected as significantly fewer tumour cells were demonstrated, but still to be categorized as CPS≥10 and in the right bin, due to a relatively robust margin from the cut-off point.
No plausible reason as e.g. lot no. of the primary antibodies causing the general reduced analytical sensitivity and accuracy for the two SP263 IHC assays could be identified.
The observed reduced analytical sensitivity of the SP263 assay for both TPS (NSCLCs) and CPS (TNBC) is as mentioned in contrast to the expected performance and previously published data. The recently published comparative study by Noske et al.[1] "Interassay and interobserver comparability study of four programmed death-ligand 1 (PD-L1) immunohistochemistry assays in triple-negative breast cancer" indicate an enhanced proportion of PD-L1 positive cases using SP263 compared to 22C3 using CPS≥10 as cut-off. Same has been observed in several publications focusing on PD-L1 TPS in NSCLC as e.g. Tsao et al.[2], Torlakovic et al.[3] The reduced analytical sensitivity also being in contrast to the data recently published by Sompuram et al.[4] showing that "the SP263 assay was the most sensitive PD-L1 approved assay" when comparing the levels for low limit of PD-L1 demonstration in IHC calibrators containing different levels and dynamic range of purified protein PD-L1 analytes.
Laboratories obtaining an insufficient score are recommended to continue to use the two SP263 based PD-L1 assays with vendor guided protocol settings, as they historically in the NordiQC assessments have generated high qualitative results, but also highly encouraged to perform in-house metrics of the PD-L1 results obtained to monitor and document these and hereby verify the proportion of positive and negative results are on par to levels expected and published for the cancer types in question.

The Dako/Agilent 22C3 pharmDx assay GE006 for Dako Omnis was used by 18% of the participants and in this assessment the most successful assay providing a pass rate of 97% (81% optimal) when applied by protocol settings in compliance with vendor recommendations (see Table 3).
Similar to the data generated in previous runs, it was observed that the PD-L1 22C3 GE006 assay for Omnis was more successful compared to the 22C3 pharmDx SK006 for Autostainer Link 48. Cumulated data for the latest 6 successive runs has shown a pass rate of 99% (115 of 116) for laboratories using GE006 by vendor recommended protocol settings. In comparison a pass rate of 83% 101 of 121) for laboratories using SK006 by vendor recommended protocol settings 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 latest six successive runs performed. The superior performance of GE006 might in part be related to a more consistent reproducibility of the 22C3 pharmDx assay on the fully automated Dako Omnis platform compared to the assay when applied on the semi-automated Autostainer Link 48
In this context it has to be emphasized that the 22C3 GE006 assay for Dako Omnis is by Dako/Agilent only validated for PD-L1 status and predictive claim in NSCLC with TPS as scoring system and at present not validated by Dako/Agilent for any indication with CPS as scoring system including TNBC.

The Dako/Agilent 22C3 pharmDx assay SK006 for Autostainer Link 48 was used by 16% of the participants and provided a pass rate of 83% (58% optimal) when applied by protocol settings in compliance with vendor recommendations (see Table 3). The 22C3 SK006 assay was frequently (n=11) applied off-label on a non-Autostainer Link 48 platform as e.g. BenchMark Ultra/GX/XT (Ventana/Roche) and as shown in table 2 with inferior performance. In total 7 laboratories used the SK006 assay on BenchMark with a pass rate of 57%, 14% optimal.

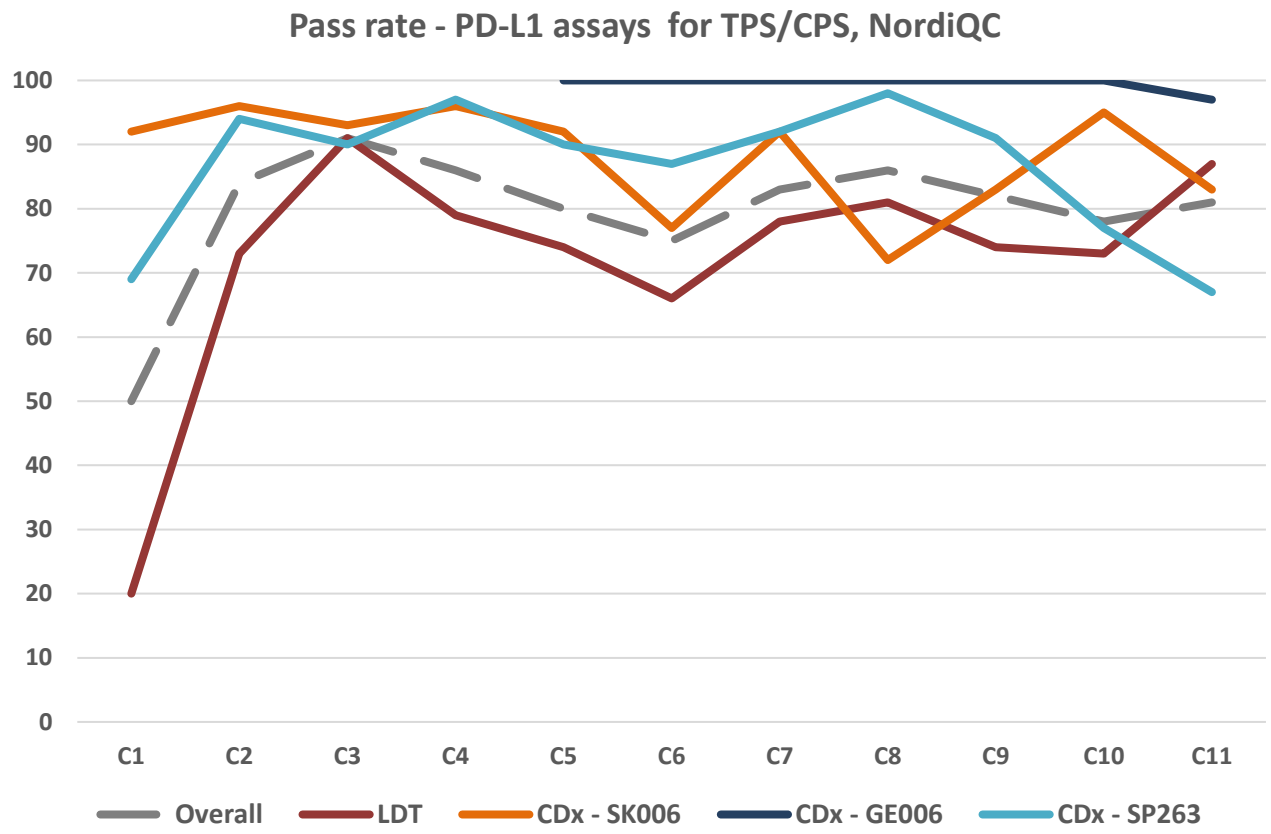
The Dako/Agilent pharmDx SK005 28-8 for Autostainer Link 48 was used by two laboratories. Both used the recommended protocol settings and both results being assessed as sufficient (Good).

Laboratory developed (LD) assays either based on a concentrated Ab, a RTU Ab without any predictive claim or a companion diagnostic assay not used strictly accordingly to vendor recommendations were applied by 37% (83 of 225) of the participants, which was reduced compared to 50% in the previous

assessment. For this group a pass rate of 87% was observed being a significant improvement compared to the level seen in the latest assessment runs. Focusing on the performance of PD-L1 LD assays from C2-C11, excluding the initial run C1 and start-up phase to identify "best practice LD assays", the mean pass rate for LD assays has been 78% (range 66%-91%) compared to e.g. 99% for the 22C3 GE006 pharmDx (Dako/Agilent) and 88% for both the SP263 assay (Ventana/Roche) and 22C3 SK006 pharmDx (Dako/Agilent).

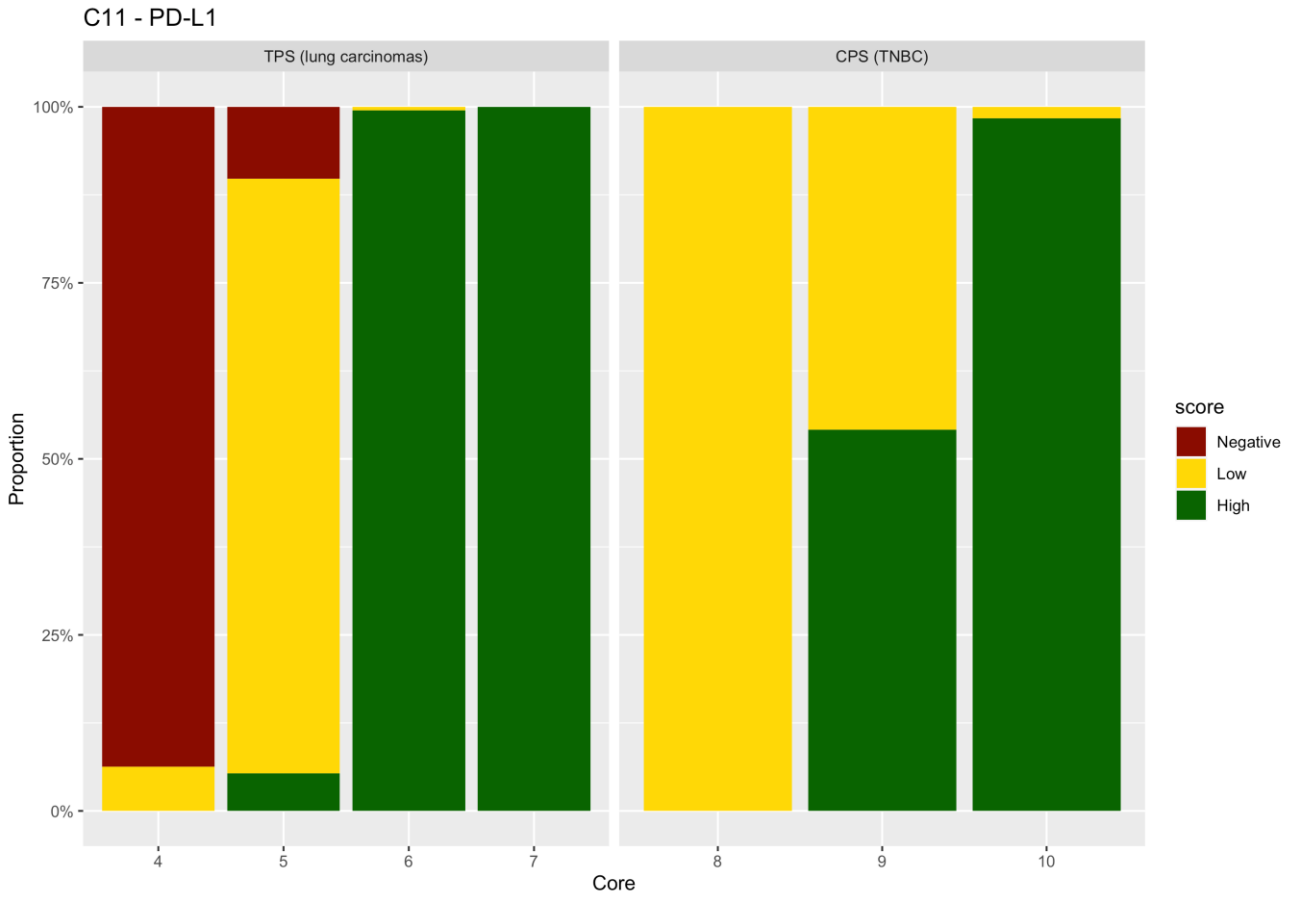
The performance of most commonly used IHC CDx and LD assays for PD-L1 is summarized and shown in Graph 3 below.

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



The mAb clone 22C3 was the most widely used concentrated Ab within a LD assay (n=39) providing an "all-time" high pass rate of 95%, 49% 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) and these now seem to be widely consolidated within the laboratories providing a pass rate fully comparable and even superior to some of the companion diagnostic assays in this assessment as show in Graph 3 above.
As mentioned in previous reports the performance of mAb clone 22C3 on BOND III / BOND MAX (Leica Biosystems) has shown to be inferior. Cumulated data for runs C8 - C11 focusing on the performance of mAb clone 22C3 on the BOND platforms have shown a pass rate of 29% (4 of 14), 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, but as mentioned same trend have now been observed in 4 successive runs. For the BOND platform, the rmAb CAL10 as concentrated format reported in run C10 and the mAb clone E1L3N by protocol settings as described above seem to be preferable for the development and validation of a LD PD-L1 protocol with focus on TPS/CPS scoring.

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 TNBCs.



Graph 4. **NordiQC PD-L1 run C11: Tumour Proportion scores (TPS) in NSCLCs (core no. 4-7) and Combined Positive Score (CPS) in TNBCs (core no. 8-10).**

As seen in Graph 4, a relatively high consensus rates were observed for the tissue core 4, 6, 7, 8 and 10, whereas the consensus rates were significantly lower in tissue core 9 but also reduced in core 5. This also correlates with the insufficient and false negative results typically being seen in these two tissue cores.

Controls
Throughout all assessments for PD-L1 TPS/CPS tonsil and placenta have been used as positive and negative tissue controls and tonsil has been found to be superior to placenta, as tonsil typically display a dynamic and clinically relevant range of PD-L1 expression levels from weak, low to high, whereas placenta typically only contain cells (throphoblasts) with high level PD-L1 expression.
In tonsil, protocols with optimal results for PD-L1 TPS/CPS status typically provide 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 intra- and interfollicular lymphocytes and macrophages showing a coarse punctuated granular cytoplasmic staining reaction. No staining reaction in the vast majority of lymphocytes and normal stratified squamous epithelial cells.
It has been observed that different assays and/or clones for PD-L1 TPS/CPS status do give different staining patterns in tonsil, which must be taken into account when evaluating the reaction pattern and to verify if the result is as expected. The rmAb clone SP263 (741-4905, 790-4905/4907, Ventana/Roche) typically provide 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 and E1L3N typically a stronger staining reaction in more germinal centre macrophages were seen compared to mAb clone 22C3, when the clones still 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.

In contrast to previous assessments, placenta was found to be a relatively valuable tool to evaluate the level of analytical/technical sensitivity of the PD-L1 IHC test. In the protocols, irrespectively of clone applied, providing an optimal result, both the membrane and cytoplasmic compartment of trophoblasts was distinctively demonstrated. For protocols providing an insufficient and false negative result in one of the carcinomas, only the membrane was clearly demonstrated and no or only a faint cytoplasmic staining reaction in the trophoblasts was observed, see photos 2a and 2b.

1. Noske A, Wagner DC, Schwamborn K, et al. Interassay and interobserver comparability study of four programmed death-ligand 1 (PD-L1) immunohistochemistry assays in triple-negative breast cancer. Breast. 2021;60:238-244.

2. Ming Sound Tsao, Keith M. Kerr, Mark Kockx, et al. PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples: Results of Blueprint Phase 2 Project. Journal of Thoracic Oncology. 2018;13(9):1302-1311

3. Torlakovic E, Lim HJ, Adam J, et al. "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod Pathol. 2020;33(1):4-17.

4. Sompuram, Seshi R et al. Quantitative comparison of PD-L1 IHC assays against NIST standard reference material 1934." Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc. 2022;35(3): 326-332.

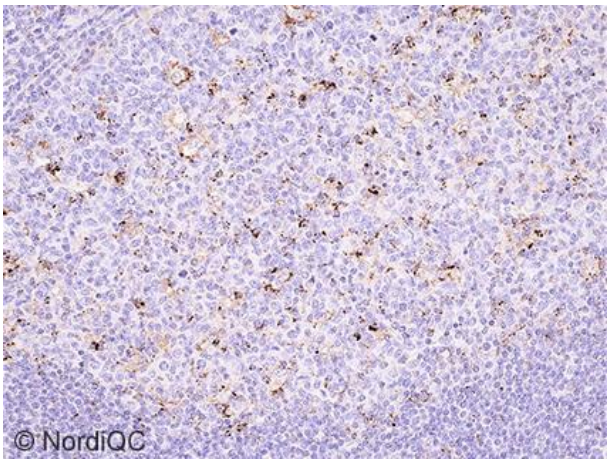


Fig. 1a
Optimal staining result of tonsil using the PD-L1 IHC 22C3 pharmDx kit, Dako/Agilent following the vendor recommended protocol settings.
A weak to moderate, but distinct punctuated membranous staining reaction of germinal centre macrophages and dispersed lymphocytes is seen.
No staining reaction is seen in the vast majority of lymphocytes.
Also compare with Figs. 2a – 6a, same protocol.

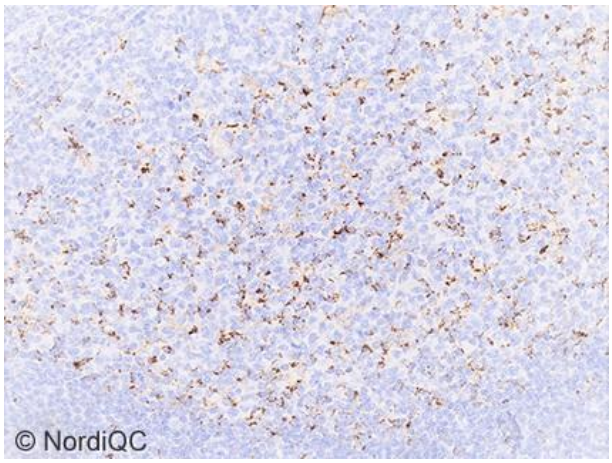


Fig. 1b
Staining result of tonsil, using the rmAb clone SP263 741-4905 Ventana/Roche on BenchMark Ultra.
The result in tonsil was fully comparable to the expected level as seen in Fig. 1a and obtained by the 22C3 pharmDx kit. However still an insufficient result was seen in three of the included carcinomas in the TMA, as shown in Figs. 3b – 5b.
The protocol as such provided an overall insufficient result characterized by a too low level of analytical sensitivity. In this context is was concluded that tonsil isolated cannot be used as quality control for PD-L1 IHC testing.

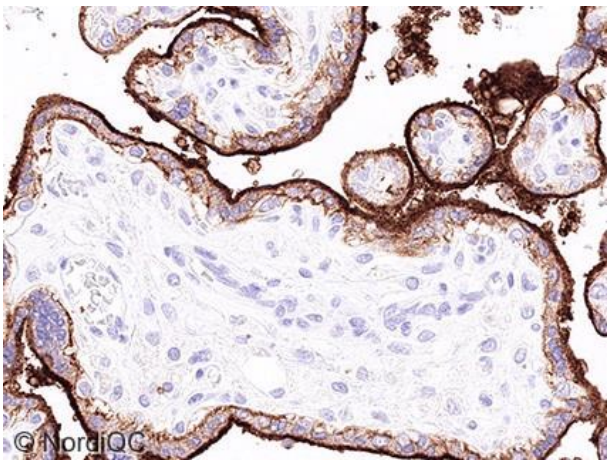


Fig. 2a
Optimal staining result of the placenta, using the same protocol as in Fig. 1a.
Virtually all trophoblasts show a strong membranous staining reaction, but also a weak to moderate staining reaction of the cytoplasmic compartment.

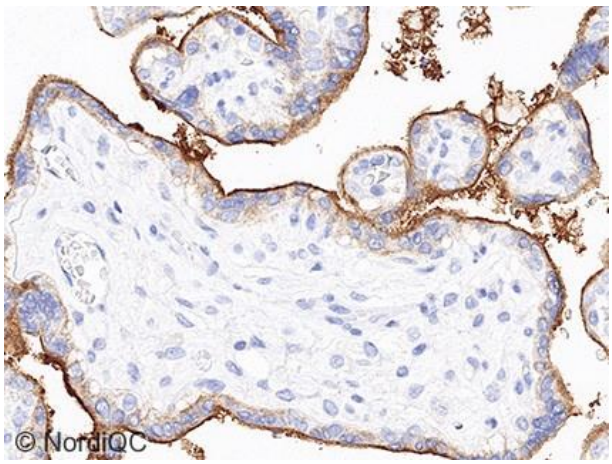


Fig. 2b
Staining result of the placenta, using the same protocol as in Fig. 1b.
The vast majority of trophoblasts show a moderate membranous staining reaction, whereas the cytoplasm only display a barely perceptible staining reaction.
In this assessment placenta was found to be a valuable supplemental tool to tonsil for the evaluation of level of analytical/technical sensitivity of the PD-L1 IHC test when focusing on the staining pattern in the trophpoblasts.
Also compare with Figs. 3b - 5b, same protocol.

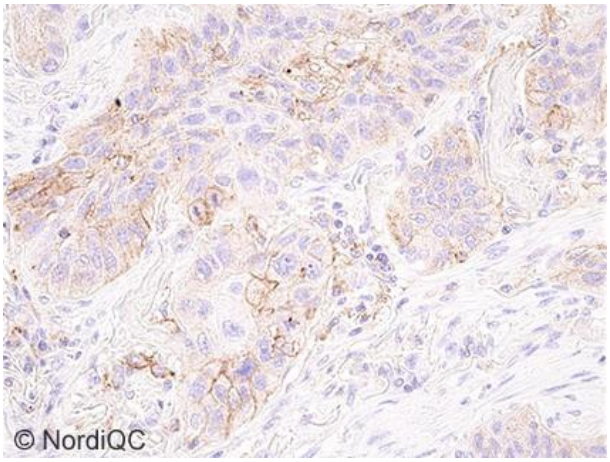


Fig. 3a
Optimal staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1a and 2a.
A weak to moderate, but distinct staining reaction is seen in dispersed tumour cells.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA®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

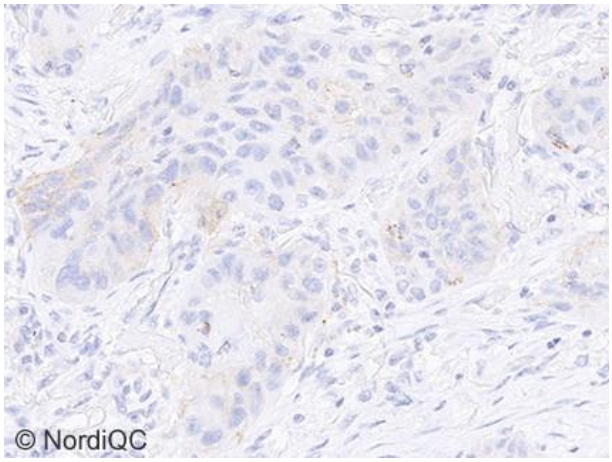


Fig. 3b
Insufficient staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1b and 2b.
Virtually no staining reaction in the tumour cells is seen and the PD-L1 category being changed from the expected TPS low to negative and the tumour not being eligible for immune therapy.
Compare to the expected result as shown in Fig. 3a.

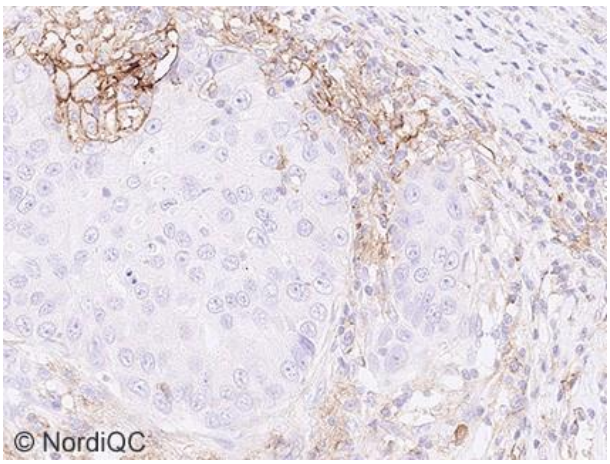


Fig. 4a
Optimal staining result of the TNBC, tissue core no. 9, using the same protocol as in Figs. 1a - 3a.
A weak to strong and distinct staining reaction is seen in the immune cells in the vicinity of tumour cells.
The tumour was categorized as CPS≥10 and thus eligible for immune therapy with KEYTRUDA®.

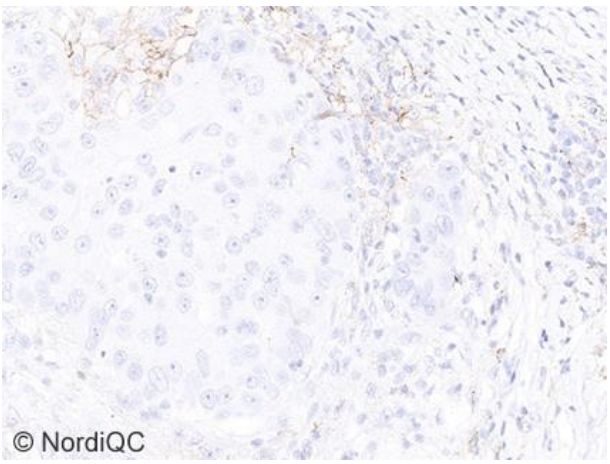


Fig. 4b
Insufficient staining result of the TNBC, tissue core no. 9, using the same protocol as in Figs. 1b – 3b.
Only dispersed immune cells are demonstrated and significantly reduced compared to the level expected and overall the result categorized as CPS<10. PD-L1 and category being changed from the expected TPS low to negative and the tumour not being eligible for immune therapy.
Compare to the expected result as shown in Fig. 4a.

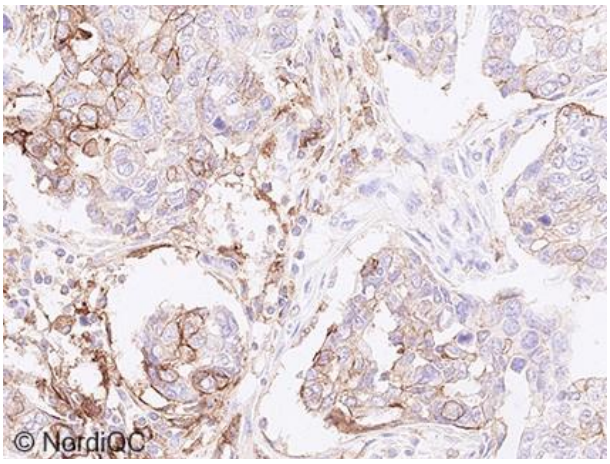


Fig. 5a
Optimal staining result of the TNBC, tissue core no. 10, using the same protocol as in Figs. 1a - 4a.
A weak to strong and distinct staining reaction is primarily seen in the tumour cells but also in scattered immune cells.
The tumour was categorized as CPS≥10 and thus eligible for immune therapy with KEYTRUDA®.

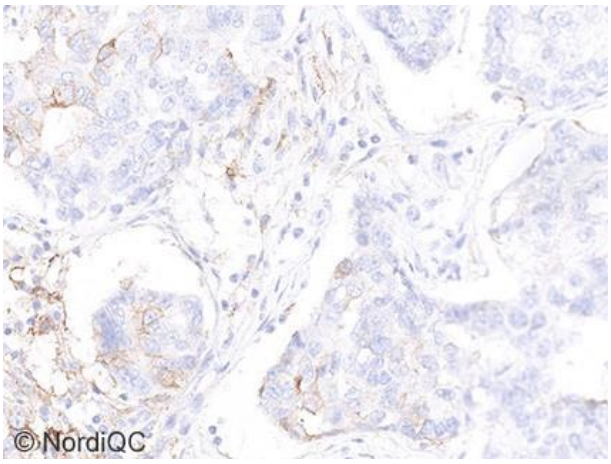


Fig. 5b
Insufficient staining result of the TNBC, tissue core no. 10, using the same protocol as in Figs. 1b – 4b.
Only few tumour and immune cells are demonstrated and the number and intensity being significantly reduced compared to the level expected.
Compare to the expected result as shown in Fig. 5a.

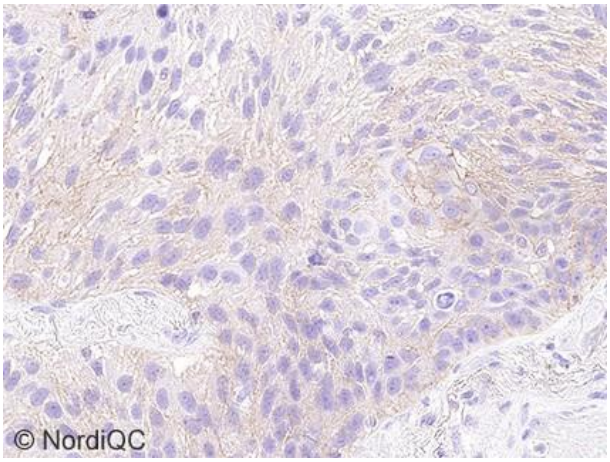


Fig. 6a
Optimal staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1a - 5a based on the 22C3 pharmDx, Agilent/ Dako following the vendor recommended protocol settings.
A weak to moderate, but distinct staining reaction is seen in dispersed tumour cells.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA® (different regional cut-offs occur).

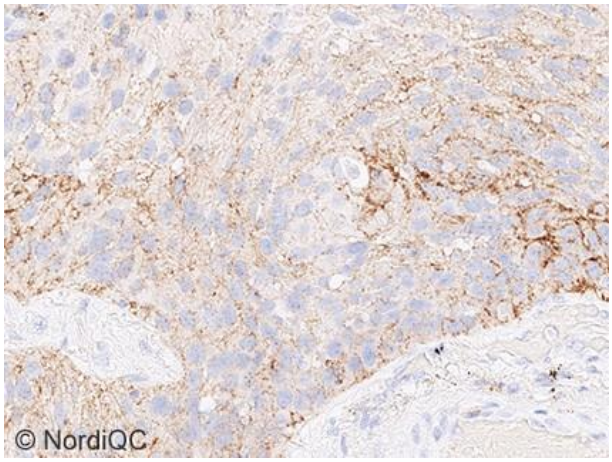


Fig. 6b
Insufficient staining result of the NSCLC, tissue core no. 5, using the 22C3 pharmDx. Agilent/Dako by laboratory modified protocol settings.
A granular and extended membranous staining reaction for PD-L1 is seen in most tumour cells changing the PD-L1 category from TPS low (≥1-49%) to TPS high (≥50%).
The protocol was based on a detection system with tyramide amplification.

SN/LE/RR 12.06.22