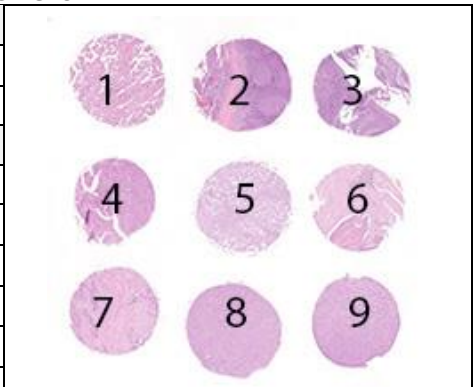
Assessment Run C10 2021
PD-L1 IC

Purpose
This assessment in the NordiQC Companion module of PD-L1 IC primarily focused on evaluation of the analytical accuracy of the PD-L1 IHC assays performed by the participating laboratories to identify patients with urothelial carcinomas or triple negative breast carcinomas (TNBC) to be treated with TECENTRIQ® as immune therapy. The PD-L1 SP142 IHC assay (741-4860, Ventana/Roche) was used as reference standard method. Accuracy was evaluated in six carcinomas with the dynamic and critical relevant expression levels of PD-L1 characterized by tumour-infiltrating immune cell score (IC). The assessment mark obtained 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 IC C10 assessment**

Tissue controls	**PD-L1 IHC reaction pattern**
1. Placenta	See control section
2-3. Tonsil	See control section
Carcinomas	**IC score***
4. Urothelial carcinoma	<5% IC
5. Urothelial carcinoma	≥5% (IC, 5-10%)
6. Urothelial carcinoma	≥5% (IC, 5-10%)
7. TNBC**	<1% IC
8. TNBC	≥1% (IC, 1-5%)
9. TNBC	≥1% (IC, 1-10%)



* Tumour-infiltrating immune cell score (IC) determined by PD-L1 SP142 IHC (741-4860, Ventana/Roche) performed in NordiQC reference lab.
** Triple negative breast carcinoma.

All tissues were fixed in 10% neutral buffered formalin.

The participating laboratories were asked to perform the PD-L1 IHC assay accordingly to the protocol used in the laboratory, evaluate the PD-L1 expression level using IC score as read-out method and submit these scores to NordiQC.
This allowed both an assessment of the technical performance / analytical accuracy of the PD-L1 IHC assays but also information on the reproducibility and concordance of the PD-L1 expression read-out results among the laboratories.

PD-L1 IC IHC, 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. Every twenty-fifth slide was thus stained for PD-L1 using the CE IVD / FDA approved PD-L1 SP142 IHC assay (741-4860, Ventana/Roche). During the assessment, IC categories for each tissue core on the submitted slides were compared to the level in the nearest reference slide of PD-L1 (SP142).

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>IC score 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 in all of the included tissues.
The PD-L1 expression in one or more tissues varies significantly from the expected IC scores, but still in right category.
The protocol may be optimized to ensure analytical accuracy and/or improved counter staining, morphology and signal-to-noise ratio.
<u>IC score is concordant to the NordiQC reference data in all carcinomas.</u>

Criteria for assessing a staining as <u>Borderline</u> include:
The staining is considered insufficient, e.g., because of a generally too weak staining, a false negative staining or a false positive staining reaction in one of the included tissues. The protocol should be optimized.
<u>IC score is **not** found 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 staining in more than one of the included tissues.
An optimization of the protocol is urgently needed.
<u>IC score is **not** found 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 reaction in non-immune cells hampering the interpretation.

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

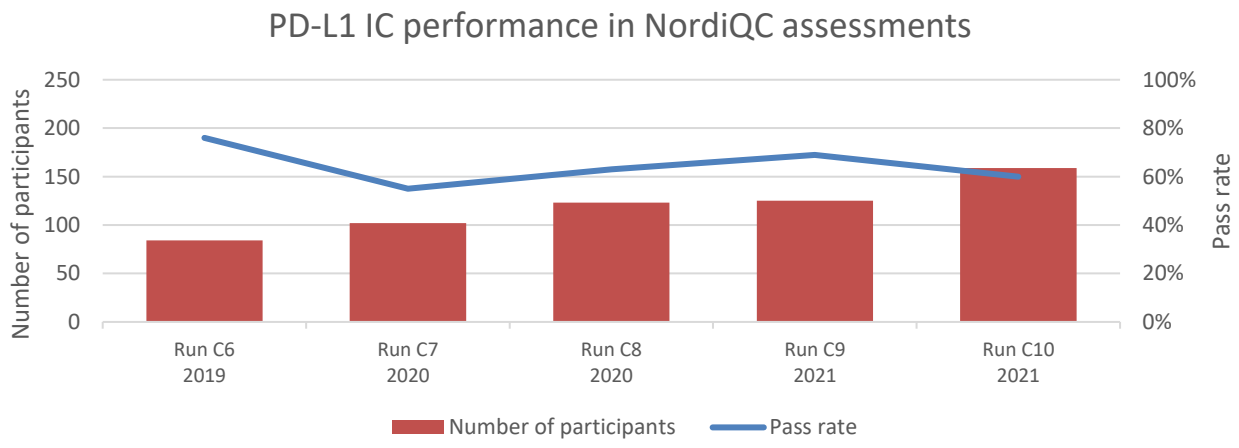
Participation

Number of laboratories registered for PD-L1 IC IHC C10	193
Number of laboratories returning PD-L1 IC IHC	159 (82%)
Number of laboratories returning PD-L1 scoring sheet	158

Results: 159 laboratories participated in this assessment and 60% 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 received advice if the result being insufficient but were not included in this report.

Performance history
This was the fifth NordiQC assessment of PD-L1 IC. The overall pass rate decreased compared to the results obtained in C8 and C9 (see Graph 1). The number of new participants have increased significantly in this run C10, from 125 in C9 till 159.

Graph 1. **Proportion of sufficient results for PD-L1 IC in the NordiQC runs performed**



Conclusion
This was the fifth NordiQC assessment of PD-L1 for IC in urothelial carcinoma and TNBC in the companion module. 159 laboratories participated and a relatively low pass rate of 60% was observed.
The PD-L1 SP142 companion diagnostic (CDx) IHC assay product no. 741-4860 and the IHC assay 790-4860 both from Ventana/Roche were the most successful assays for the evaluation of PD-L1 status in urothelial carcinomas and TNBCs to guide treatment with TECENTRIQ® as immune therapy providing a pass rate of 78% and 76%, respectively, if following the vendor recommended protocol settings. Other PD-L1 CDx assays as SP263 (741-4905, Ventana/Roche) and 22C3 (SK006/GE006, Dako/Agilent) being very successful in the NordiQC PD-L1 TPS/CPS assessments provided no sufficient staining results. The

insufficient results were typically characterized by an either too weak or completely false negative staining reaction in immune cells or a strong staining reaction in tumour cells in one or more of the carcinomas compromising the evaluation of PD-L1 reaction in immune cells – most likely because these protocols have been developed and calibrated to primarily demonstrate PD-L1 expression in tumour cells and to imitate the performance of the Dako/Agilent SK006 pharmDx 22C3 PD-L1 assay for treatment with KEYTRUDA®.

Table 2. **Assessment marks for IHC assays and antibodies run C10, PD-L1 IC**

CE-IVD / FDA approved PD-L1 assays	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP142, **741-4860**[3]	73	Ventana/Roche	25	32	9	7	78%	35%
rmAb clone SP142, **741-4860**[4]	1	Ventana/Roche	0	0	1	0	-	-
rmAb clone SP263, **741-4905**[3]	5	Ventana/Roche	0	0	5	0	0%	0%
mAb clone 22C3 pharmDX, **SK006**	3	Dako/Agilent	0	0	2	1	-	-
mAb clone 22C3 pharmDX, **GE006**	4	Dako/Agilent	0	0	3	1	-	-
Antibodies[7] for laboratory developed PD-L1 assays, concentrated antibodies	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
mAb clone **22C3**	4	Dako/Agilent	0	0	1	3	-	-
mAb clone **405-9A11**	1	Diagnostic BioSystems	0	0	0	1	-	-
mAb clone **E1L3N**	1	Cell Signaling	0	0	1	0	-	-
rmAb clone **CAL10**	1 1	Zytomed Biocare Medical	0	0	2	0	-	-
rmAb clone **QR001**	1	Quartett	0	0	1	0	-	-
rmAb clone **ZR3**	1	Gene Tex	0	0	1	0	-	-
Ready-To-Use antibodies[8]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP142, **790-4860 (VRPS)**[5]	21	Ventana/Roche	11	5	3	2	76%	52%
rmAb clone SP142, **790-4860 (LMPS)**[6]	38	Ventana/Roche	10	12	12	4	58%	26%
rmAb clone SP263, **790-4905**	3	Ventana/Roche	0	0	3	0	-	-
rmAb clone IHC411, **IHC424**	1	GenomeMe	0	0	1	0	-	-
Total	159		46	49	45	19		
Proportion			29%	31%	28%	12%	60%	

1) Proportion of sufficient stains (optimal or good) (≥5 assessed protocols).
2) Proportion of optimal results (≥5 assessed protocols).
3) This product has a locked protocol on all BenchMark platforms and cannot be changed.
4) RTU product applied on another platform than developed for.
5) Vendor recommended protocol settings – RTU product used in compliance to protocol settings, platform and package insert.
6) Laboratory modified protocol settings for a RTU product applied either on the vendor recommended platform(s) or other platforms.
7) mAb: mouse monoclonal antibody, rmAb: rabbit monoclonal antibody.
8) Ready-To-Use antibodies without predictive claim.

Detailed Analysis

CE IVD / FDA approved assays

SP142 (741-4860, Ventana/Roche): In total, 25 of 73 (35%) 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) for 48 min., 16 min. incubation of primary Ab and OptiView with OptiView Amplification as detection system. 57 of 73 (78%) produced a sufficient staining result (optimal or good).

Table 3 summarizes the proportion of sufficient and optimal marks for the most commonly used CDx assays with a predictive claim. 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 assays	**Vendor recommended protocol settings[1]**		**Laboratory modified protocol settings[2]**	
	Sufficient	Optimal	Sufficient	Optimal
Ventana BenchMark GX, XT, Ultra rmAb SP142, **741-4860**	57/73 (78%)	25/73 (35%)	-	-
Ventana BenchMark GX, XT, Ultra rmAb SP263, **741-4905**	0/5	0/5	-	-
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	-	-	0/1	0/1
Dako Omnis mAb 22C3 pharmDX, **GE006**	0/3	0/3	-	-

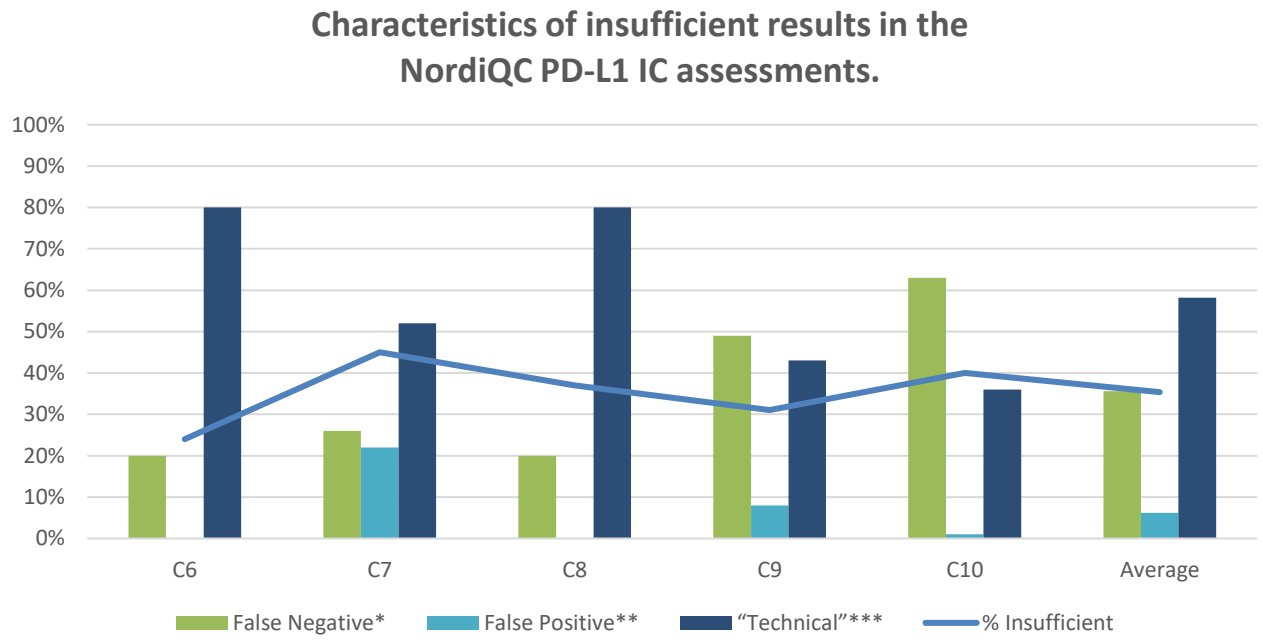
1) Protocol settings recommended by vendor – Retrieval method and duration, Ab incubation times, detection kit, IHC stainer/equipment.
2) Modifications in one or more of parameters mentioned above. Only protocols performed on the specified vendor IHC stainer are included.

Ready-To-Use antibodies for laboratory developed (LD) assays
SP142 (790-4860, Ventana/Roche): In total, 21 of 57 (37%) protocols were assessed as optimal. Protocols with optimal results were typically based on HIER in CC1 (efficient heating time 32-64 min.), 16-32 min. incubation of primary Ab and OptiView with OptiView Amplification as detection system. Using these settings, 38 of 57 (67%) produced a sufficient staining result.

Comments – accuracy of PD-L1 IHC using IC scoring to guide treatment with TECENTRIQ®
In this fifth NordiQC run for PD-L1 IC in the companion module C10, a pass rate of 60% was observed for the participants performing PD-L1 IHC assays to identify patients with urothelial carcinomas and triple negative breast carcinomas (TNBC) to be treated with TECENTRIQ® as immune therapy using the tumour-infiltrating immune cell score (IC) as scoring method.
The pass rate had decreased compared to the result obtained in run C9. In this run C10, the number of participants using the PD-L1 IHC assay based on rmAb clone SP142 from Ventana/Roche increased significantly to 84% (133 of 159) compared to 75% (94 of 125) and 73% (90 of 123) in run C9 and C8, respectively. Only protocols based on the rmAb clone SP142 obtained sufficient staining results.
The central parameters potentially affecting pass rates in IHC proficiency schemes were identical in all the five NordiQC assessment runs for PD-L1 IC. Of critical importance, the same assessment criteria, reference standard methods and scoring guidelines were applied. The materials / carcinomas selected and used for the individual assessment runs were different and the decreased pass rate might be caused by more challenging material circulated in this run.

It was observed that insufficient results were most frequently characterized by a reduced proportion of cells demonstrated or a completely false negative staining reaction of immune cells in one or more of the tissue cores and was seen in 63% (40 of 64) of the insufficient results. In 36% (23 of 64) of the insufficient results, a reduced proportion and/or too weak specific staining reaction of immune cells combined with an excessive staining reaction of tumour cells compromising the scoring and PD-L1 status in the immune cells. One insufficient staining result was caused by an increased proportion of immune cells demonstrated giving a false positive staining reaction. Graph 2 shows the main characteristics of insufficient results in the five NordiQC PD-L1 IC runs performed.

Graph 2. **Characteristics of insufficient results in the five NordiQC PD-L1 IC runs.**



* IC score change from positive to negative in one or more of the included carcinomas.
** IC score change from negative to positive in one or more of the included carcinomas.
*** Interpretation compromised e.g. by poor-signal-to noise ratio, poor morphology, excessive cytoplasmic staining reaction etc.

The Ventana/Roche PD-L1 SP142 assay 741-4860 with predictive claim for TECENTRIQ® was used by 46% of the participants and provided a pass rate of 78% (57 of 73) when applying protocol settings in compliance with the vendor recommendations. The assay is locked for central protocol settings and based on HIER in CC1 for 48 min., incubation in primary Ab for 16 min. (Ultra/XT/GX) and use of OptiView with Amplification as detection system. Despite the locked protocol conditions for the assay, some laboratories submitted protocols with reported modified settings indicating change in efficient heating time of HIER, primary Ab and other detection system applied – e.g. UltraView and OptiView without Amplification. The various protocol settings submitted were disregarded for the assay product no. 741-4860 in this report and all protocols thus compiled as used by vendor recommended protocol settings as shown in Tables 2 and 3.

The Ventana/Roche PD-L1 SP142 assay 790-4860 without any predictive claim and available as an analytical or generic PD-L1 assay was used by 36% of the participants. This assay is based on same recommended protocol settings as the CDx products 741-4860, but with ordinary options for laboratories to modify the protocol settings in their optimization and validation process for the implementation of the test. Overall, the SP142 790-4860 format gave a an almost identical pass rate and proportion of optimal results, when using the vendor recommended protocol settings, compared to the corresponding CDx format of the same clone as seen in Table 2. If modifying the protocol, a reduced pass rate was seen.

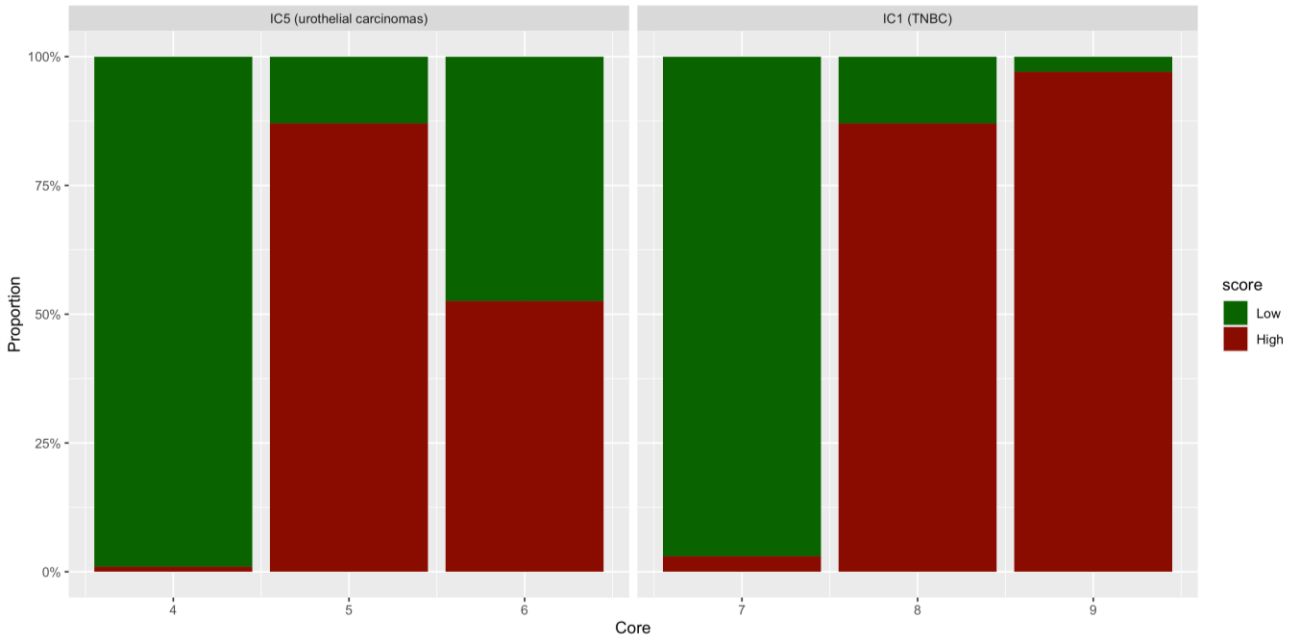
For both Ventana/Roche PD-L1 SP142 assays, a significantly reduced pass rate was seen in this run C10 compared to run C9. In run C9 pass rates at 91% and 93% were obtained when using the vendor recommended protocol settings, compared to 78% and 71% in this run C10. No plausible reason for the general reduced analytical sensitivity and accuracy for the two SP142 IHC assays could be identified. In this run C10, the number of participants increased significantly. 52 new laboratories participated for the first time, of which 83% (43 of 52) used one of the two Ventana/Roche SP142 assays with an overall pass rate of 67% (29 of 43). Laboratories that participated in both run C9 and C10 and used one of the SP142 assays, obtained an overall pass rate at 73% (66 of 90). The pass rate for both "new" and "old" participants is still significantly lower, compared to the level seen in the previous run C9. The material circulated for this run C10 might have been more challenging compared to the previous runs, but still a relatively high pass rate of 78% for the locked CDx assay 741-4860 was obtained.
At this point it also has to be emphasized that despite tonsil is the recommended and at present most reliable positive and negative tissue control with expected test performance characteristics and reaction pattern for quality control (QC), this might be challenging in real life QC. The challenges primarily related to a binary strongly positive or negative staining reaction of immune cells and epithelial cells in the tonsil, with no cells identified with low expression levels to be used as critical controls to monitor the low limit of

PD-L1 demonstration. Without such tool, the ability to evaluate the analytical precision of the PD-L1 IHC test is hampered and e.g. difficult to identify if a fluctuation of the IHC test system for PD-L1 occurs.

"Non-SP142" companion diagnostic assays as SP263 (Ventana/Roche), 22C3 pharmDx (Dako/Agilent) laboratory developed (LD) tests based on either concentrated primary Abs or RTU formats gave an overall significantly inferior performance and reduced pass rate at 0% (0 of 29) compared to the SP142 assays from Ventana/Roche used on the Ventana BenchMark platforms. The vast majority of the insufficient results were characterized by either an extensive staining reaction of tumour cells compromising the scoring of PD-L1 expression in immune cells or a too weak/negative staining reaction of immune cells. Similar observations were seen in runs C6-C9, and these data indicate a challenge for the interchangeability of the Ventana SP142 assays with other PD-L1 companion diagnostic assays and LD assays most likely designed and developed to primarily provide a staining pattern as characterized by e.g. the Dako/Agilent 22C3 pharmDx assays. One of the most influencing causes for the inferior performance of "non-SP142" assays seems to be related to the detection system applied for the Ventana SP142 assays being based on OptiView with Amplification kit (tyramide based) and the calibration of the SP142 antibody in the Ventana/Roche assays provides a performance that intensifies demonstration of immune cells and reduces staining of tumour cells.
This consideration and conclusion is fully in line with the publication of Kelly A. Schatts et al (Optimal Evaluation of Programmed Death Ligand-1 on Tumor Cells Versus Immune Cells Requires Different Detection Methods, Arch Pathol Lab Med. 2018 Aug;142(8):982-991) stressing that *"diverse sensitivities caused by the choice of the detection method should be taken into consideration when selecting PD-L1 kits or developing PD-L1 IHC laboratory-developed tests.".* Only by using the same detection system OptiView + Amplification, the classical clones as 22C3 and 28-8 could provide staining patterns largely comparable to the Ventana/Roche SP142 assays. In general, a PD-L1 IHC test must be fit-for-purpose aligning treatment, indication, scoring system and PD-L1 IHC assay.
When using alternative companion diagnostic assays or LD assays, it is strongly recommended to compare and validate these with the original assay.
The meta-analysis for PD-L1 accuracy by Torlakovic et al; "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Modern Pathology (2020) 33:4–17 also indicates that in-house or laboratory developed PD-L1 IHC assays must be developed and validated against the reference standard and approved companion diagnostic assay.

In this NordiQC PD-L1 IHC segment for IC score, the SP142 CDx assay is used as reference standard method using the associated approved read-out criteria. The results of the participants are compared directly one-to-one to the reference levels. The assessment marks only address the analytical concordance using the approved cut-off and read-out criteria focusing on IC score and e.g. application of alternative scoring systems and cut-off's for non-SP142 CDx assays are not included to adjust any option for interchangeability.

PD-L1 scoring
Participants were asked to evaluate the IC score in each of the four urothelial carcinomas (IC with 5% cut-off) and four TNBC (IC with 1% cut-off) included in the assessment material. The overall interpretation of the PD-L1 expression among the participants is shown in Graph 3.



Graph 3. **NordiQC PD-L1 run C10: Interpretation of IC in three urothelial carcinomas and three TNBC.**

As seen in Graph 3, relatively high consensus rates were observed in core 4,5, 7 and 9. Incorrect scoring was most commonly observed in tumour cores which in the reference slides were classified as PD-L1 positive (PD-L1 IC ≥ 1% or 5%). This was often linked with a less successful technical result and/or an insufficient mark.

Controls

Tonsil and placenta were used as positive and negative tissue controls. In this assessment and in concordance with the official scoring guidelines from Ventana/Roche, tonsil was found to be a recommendable positive and negative tissue control and superior to placenta. However, as mentioned above the use of tonsil as QC tool to monitor the reproducibility of the PD-L1 IC test is challenged as only a binary reaction pattern of either strongly positive cells or negative cells are identified. On the contrary no cells in tonsil are identified with low expression levels to be used as a more reliable tool to identify any test fluctuation and reduced analytical sensitivity of the PD-L1 IC test. In this context, it was observed in both this and previous assessments, that placenta might be a supplemental positive tissue control. It was as such seen that a moderate to strong staining reaction in at least dispersed cytotrophoblasts in placenta, could indicate a high and expected level of analytical sensitivity for the Ventana/Roche SP142 assays based on tyramide amplification. If these cells were identified and positive with the two SP142 assays, the results in other tissues were as expected and evaluated as successful, whereas if these cells were negative a large proportion of insufficient and false negative results in the other tissues were observed. This observation must be further validated.

When tonsil is used as positive and negative tissue control following pattern must be seen; The majority of crypt epithelial cells in the tonsil should display a strong staining reaction, while a moderate to strong staining reaction should be seen in most germinal center lymphocytes, macrophages and scattered immune cells in the interfollicular regions. No staining reaction should be seen in superficial squamous epithelial cells and mantle zone B-cells. In this assessment, it was observed that a moderate staining reaction in scattered immune cells in the interfollicular region was more challenging for the participants and could only be detected with an optimal protocol.

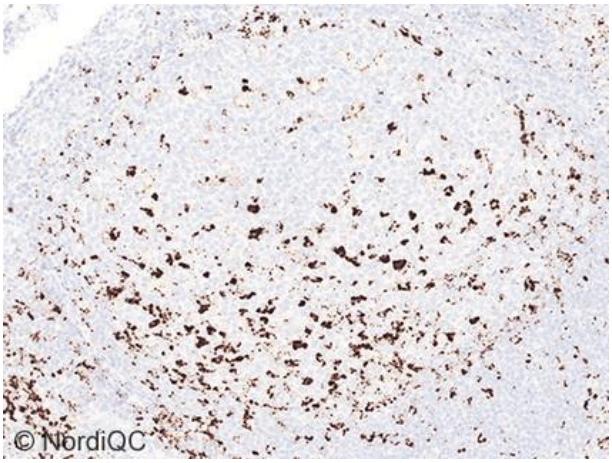


Fig. 1a
Optimal staining result of tonsil using the PD-L1 assay 741-4860 from Ventana/Roche, based on the rmAb clone SP142 following the recommended protocol settings. Same protocol used in Figs. 2a-5a.
Most germinal centre lymphocytes/macrophages and scattered interfollicular immune cells show a moderate to strong staining reaction.

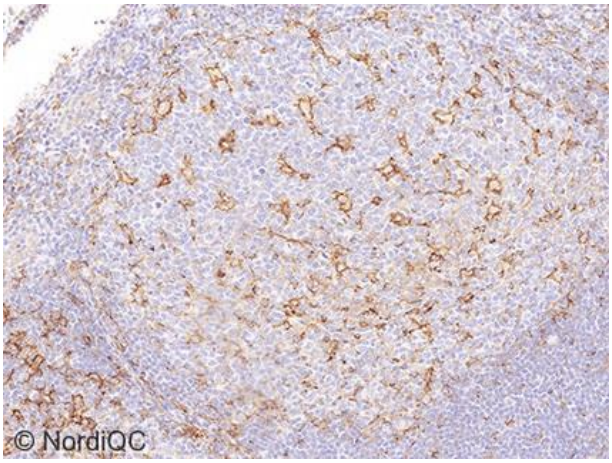


Fig. 1b
Staining result of tonsil using the rmAb clone ZR3. The protocol was based on HIER in an alkaline buffer and a 2-step detection system. Same protocol in Figs. 2b-4b.
The staining intensity of immune cells is decreased compared to the level obtained by the SP142 based assay. Compare with Fig. 1a – same area.

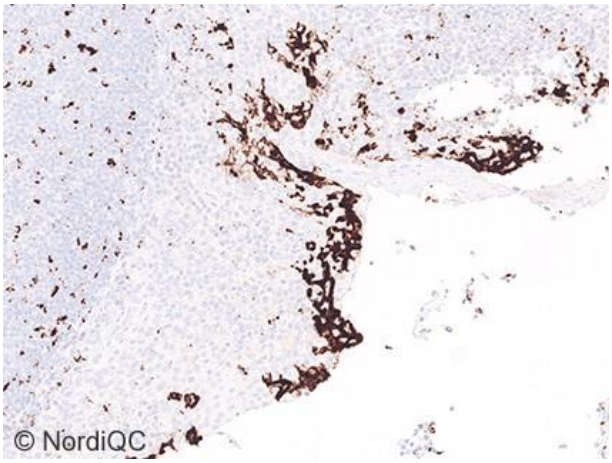


Fig. 2a
Optimal staining result of tonsil using same protocol as in Fig. 1a.
The crypt epithelial cells show an intense staining reaction, while superficial squamous epithelial cells being negative.
Also note immune cells below epithelial surface being identified.

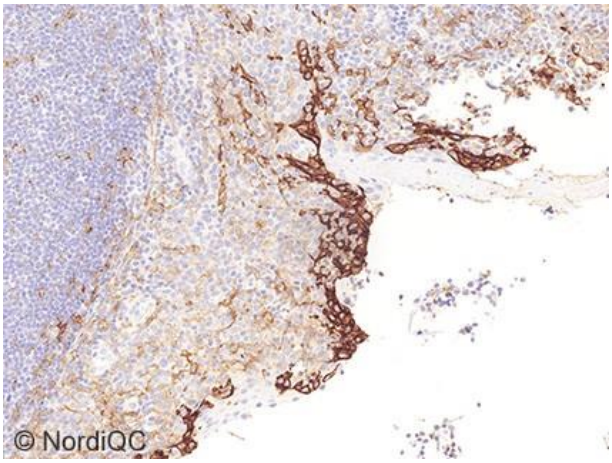


Fig. 2b
Staining result of tonsil using same protocol as in Fig. 1b.
The staining intensity of crypt epithelial cells is decreased compared to the level obtained by the SP142 based assay. A weak staining reaction is seen throughout the squamous epithelial layer. The immune cells are virtually negative.
Compare with Fig. 2a – same area.

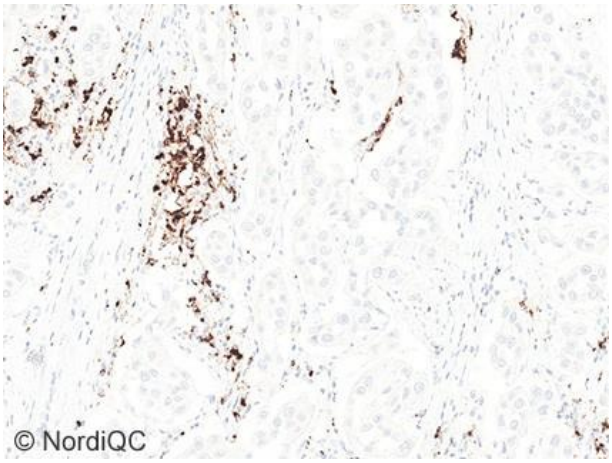


Fig. 3a
Optimal staining result of the urothelial carcinoma, tissue core no. 5, using same protocol as in Figs. 1a-2a. Virtually all tumour cells are negative and immune cells show a moderate to strong staining reaction giving an IC score of ≥5%.
The absence of staining reaction in stromal and tumour cells facilitates the evaluation of PD-L1 IC score.

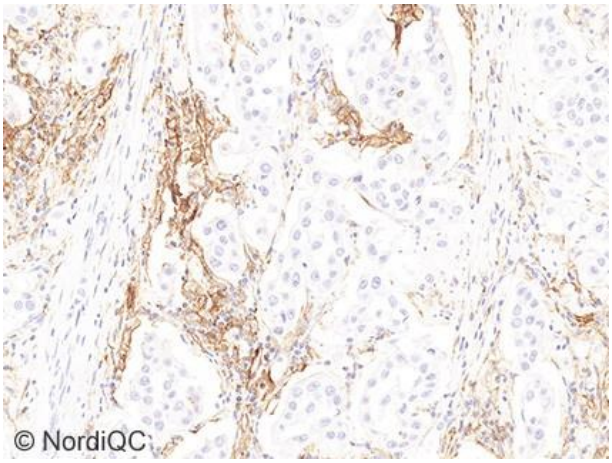


Fig. 3b
Staining result of the urothelial carcinoma, tissue core no. 5, using same protocol as in Figs. 1b-2b. Stromal cells show a moderate staining reaction, complicating the evaluation of PD-L1 expression of immune cells. Compare to the optimal result shown in Fig. 3a – same area.

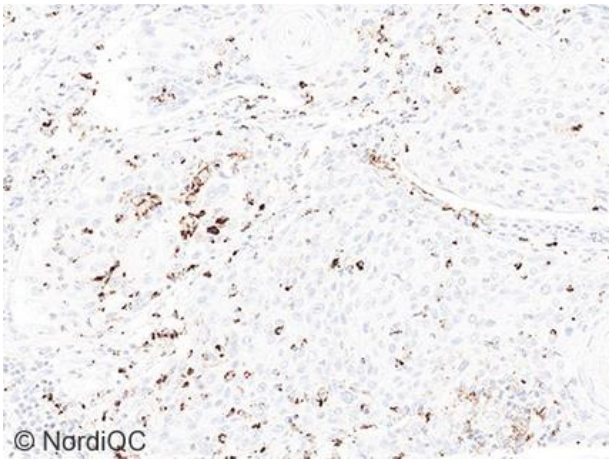


Fig. 4a
Optimal staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1a-3a. Immune cells display a moderate to strong staining reaction giving an IC score ≥5%.
The absence of staining reaction in stromal and tumour cells facilitates the evaluation of PD-L1 IC score.

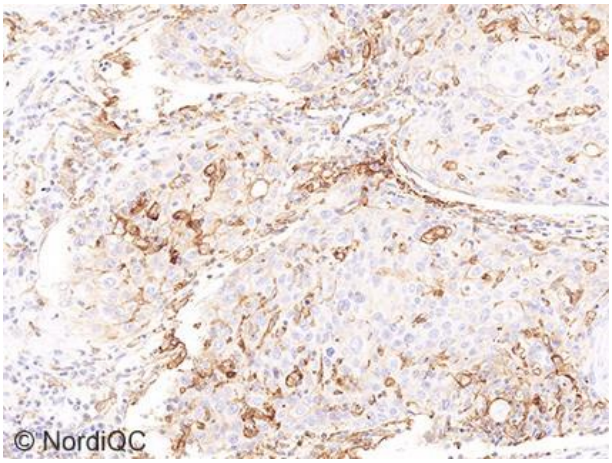


Fig. 4b
Insufficient staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1b-3b. Scattered tumour cells display a weak membranous staining reaction compromising the identification and evaluation of PD-L1 reaction in the immune cells. Compare to the optimal result shown in Fig. 4a – same area.

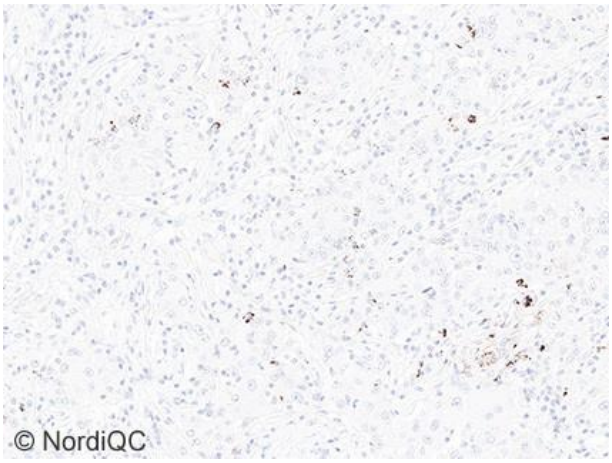


Fig. 5a
Optimal staining result of the TNBC, tissue core no. 8, using same protocol as in Figs. 1a–4a. Virtually all tumour cells are negative and immune cells show a moderate to strong staining reaction giving an IC score of ≥1%.
The absence of staining reaction in the tumour cells facilitates the evaluation of PD-L1 IC score.

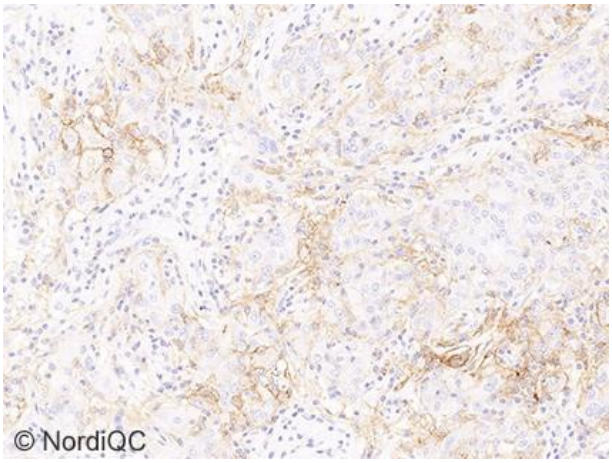


Fig. 5b
Insufficient staining result of the TNBC, tissue core no. 8, using same protocol as in Figs. 1b-4b. Tumour cells display a weak membranous staining reaction compromising the identification and evaluation of PD-L1 reaction in the immune cells. Compare to the optimal result shown in Fig. 5a – same area.

HLK/LE/SN 08.12.21