Assessment Run C8 2020

PD-L1 TECENTRIQ®

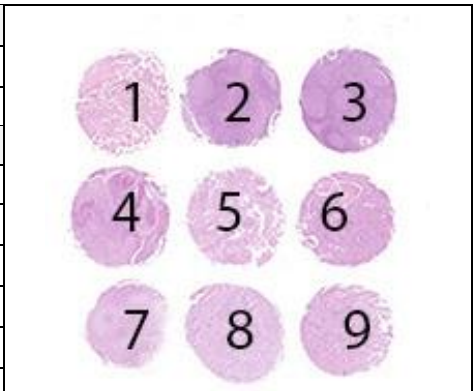
Purpose

This assessment in the NordiQC Companion module of PD-L1 TECENTRIQ® 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) 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 TECENTRIQ® C8 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
5. Urothelial carcinoma	≥5 (IC 5-10)
6. Urothelial carcinoma	≥5 (IC 5-10)
7. TNBC**	<1
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) 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 and also interpret 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 interpretation of PD-L1 expression among the laboratories.

PD-L1 TECENTRIQ® 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). 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 Borderline 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.
IC score is **not** found concordant to the NordiQC reference data in 1 of the carcinomas.

Criteria for assessing a staining as Poor 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.
IC score is **not** found concordant to the NordiQC reference data in 2 or more of the carcinomas.

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 TECENTRIQ® IHC C8	136
Number of laboratories returning PD-L1 TECENTRIQ® IHC	123 (90%)
Number of laboratories returning PD-L1 TECENTRIQ® scoring sheet	113

All slides returned after the assessment were assessed and received advice if the result being insufficient but were not included in this report.

Results: 123 laboratories participated in this assessment and 63% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Table 2 (see page 3).

Performance history
This was the third NordiQC assessment of PD-L1 for **TECENTRIQ®**. The overall pass rate increased compared to the result obtained in run C7, 2020 (see Table 3).

Table 3. **Proportion of sufficient results for PD-L1 TECENTRIQ® in the three NordiQC runs performed**

	Run C6 2019	Run C7 2020	Run C8 2020
Participants, n=	84	102	123
Sufficient results	76%	55%	63%

Conclusion
This was the third NordiQC assessment of PD-L1 for **TECENTRIQ®** in urothelial carcinoma and TNBC in the companion module. 123 laboratories participated and a relatively low pass rate of 63% was observed. The PD-L1 SP142 companion diagnostic (CDx) IHC assay product no. 741-4860 and the IHC assay 790-4860 from Ventana 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 82% and 94%, respectively. Other PD-L1 CDx assays as SP263 (741-4905, Ventana) and 22C3 (SK006/GE006, DakoAgilent) being very successful in the NordiQC PD-L1 KEYTRUDA® assessments provided no sufficient staining results. The insufficient results were typically characterized by a too strong staining reaction in tumour cells in one or more of the carcinomas compromising the interpretation 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 SK006 pharmDx 22C3 assay for treatment with KEYTRUDA®.

Table 2. **Assessment marks for IHC assays and antibodies run C8, PD-L1 TECENTRIQ® for IC scoring**

CE-IVD / FDA approved PD-L1 assays	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP142, **741-4860 (VRPS)[3]**	44	Ventana/Roche	18	18	5	3	82%	41%
rmAb clone SP142 **740-4859 (VPRS)[3]**	3	Ventana/Roche	1	2	-	-	-	-
rmAb clone SP263, **741-4905 (VRPS)[3]**	4	Ventana/Roche	-	-	4	-	-	-
rmAb clone SP263, **741-4905 (LMPS)[4]**	2	Ventana/Roche	-	-	2	-	-	-
mAb clone 22C3 pharmDX, **SK006 (VRPS)[3]**	1	Dako/Agilent	-	-	-	1	-	-
rmAb clone 28-8 pharmDX, **SK005 (VRPS)[3]**	1	Dako/Agilent	-	-	1	-	-	-
mAb clone 22C3 pharmDX, **GE006 (LMPS)[4]**	2	Dako/Agilent	-	-	1	1	-	-
Antibodies[5] for laboratory developed PD-L1 assays, concentrated antibodies	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
mAb clone **22C3**	7	Dako/Agilent	-	-	6	1	0%	0%
mAb clone **E1L3N**	2	Cell Signaling	-	-	2	-	-	-
rmAb clone **SP142**	1	Abcam	-	-	1	-	-	-
rmAb clone **CAL10**	3 1	Zytomed Biocare	-	-	4	-	-	-
rmAb clone **ZR3**	1	Zeta Corporation	-	-	-	1	-	-
Ready-To-Use antibodies[6]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP142, **790-4860 (VRPS)[3]**	18	Ventana/Roche	11	6	1	-	94%	61%
rmAb clone SP142, **790-4860 (LMPS)[4]**	25	Ventana/Roche	10	11	3	1	84%	40%
rmAb clone SP263, **790-4905 (LMPS)[4]**	3	Ventana/Roche	-	-	3	-	-	-
rmAb clone 73-10, **PA0832 (VRPS)[3]**	1	Leica Biosystems	-	-	1	-	-	-
rmAb clone MXR003**, RMA-0732**	4	Maixin	-	-	3	1	-	-
Total	123		40	37	37	9		
Proportion			33%	30%	30%	7%	63%	

1) Proportion of sufficient stains (optimal or good) (≥5 assessed protocols).
2) Proportion of optimal results (≥5 assessed protocols).
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

SP142 (741-4860, Ventana): In total, 18 of 44 (41%) 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. 36 of 44 (82%) produced a sufficient staining result (optimal or good).

SP142 (740-4859, Ventana): One protocol was assessed as optimal. The protocol was based on the recommended and locked protocol; HIER in CC1 for 48 min., 16 min. incubation of primary Ab and OptiView with OptiView Amplification as detection system. All three produced a sufficient staining result.

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**	36/44 (82%)	18/44 (41%)	-	-
Ventana BenchMark GX, XT, Ultra rmAb SP142, **740-4859**	3/3	1/3		
Ventana BenchMark XT, GX, Ultra rmAb SP263, **741-4905**	0/4	0/4	0/2	0/2
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	0/1	0/1	-	-
Dako Omnis mAb 22C3 pharmDX, **GE006**	-	-	0/2	0/2
Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005**	0/1	0/1	-	-

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): In total, 21 of 43 (49%) protocols were assessed as optimal. Protocols with optimal results were typically based on HIER in CC1 (efficient heating time 32-64 min.), 16-24 min. incubation of primary Ab and OptiView with OptiView Amplification as detection system. Using these settings, 32 of 34 (94%%) produced a sufficient staining result.

Comments – accuracy of PD-L1 IHC using IC scoring to guide treatment with TECENTRIQ®

In this third NordiQC run for PD-L1 TECENTRIQ® in the companion module C8, a pass rate of 63% 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 was increased compared to the result obtained in run C7. In this run C8, 73% (90 of 123) of the participants used the PD-L1 IHC assays based on rmAb clone SP142 from Ventana, compared to 61% (62 of 102) in run C7. 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 the two runs. Of critical importance in both run C7 and C8, the same assessment criteria, reference standard methods and scoring guidelines were applied. The materials / carcinomas selected were different in the two runs and the composition of the TMAs used in run C8 most likely was more challenging, as non-SP142 assays stained tumour cells compromising the scoring if immune cells.

It was observed that insufficient results were most frequently characterized by 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. This was observed in 80% (37 of 46) of the insufficient staining results. In 20% (9 of 46) of the insufficient results, a reduced proportion demonstrated or completely negative staining reaction of immune cells in one or more of the tissue cores was observed. Table 4 shows the main characteristics of insufficient results in the three NordiQC PD-L1 TECENTRIQ® runs.

Table 4. **Characteristics of insufficient results in the three NordiQC PD-L1 TECENTRIQ® runs.**

	False Negative*	**False Positive****	**"Technical"*****
C6 (24% insufficient)	20%	-	80%
C7 (45% insufficient)	26%	22%	52%
C8 (37% insufficient)	20%	-	80%
Average (35% insufficient)	**22%**	**7%**	**71%**

* 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.

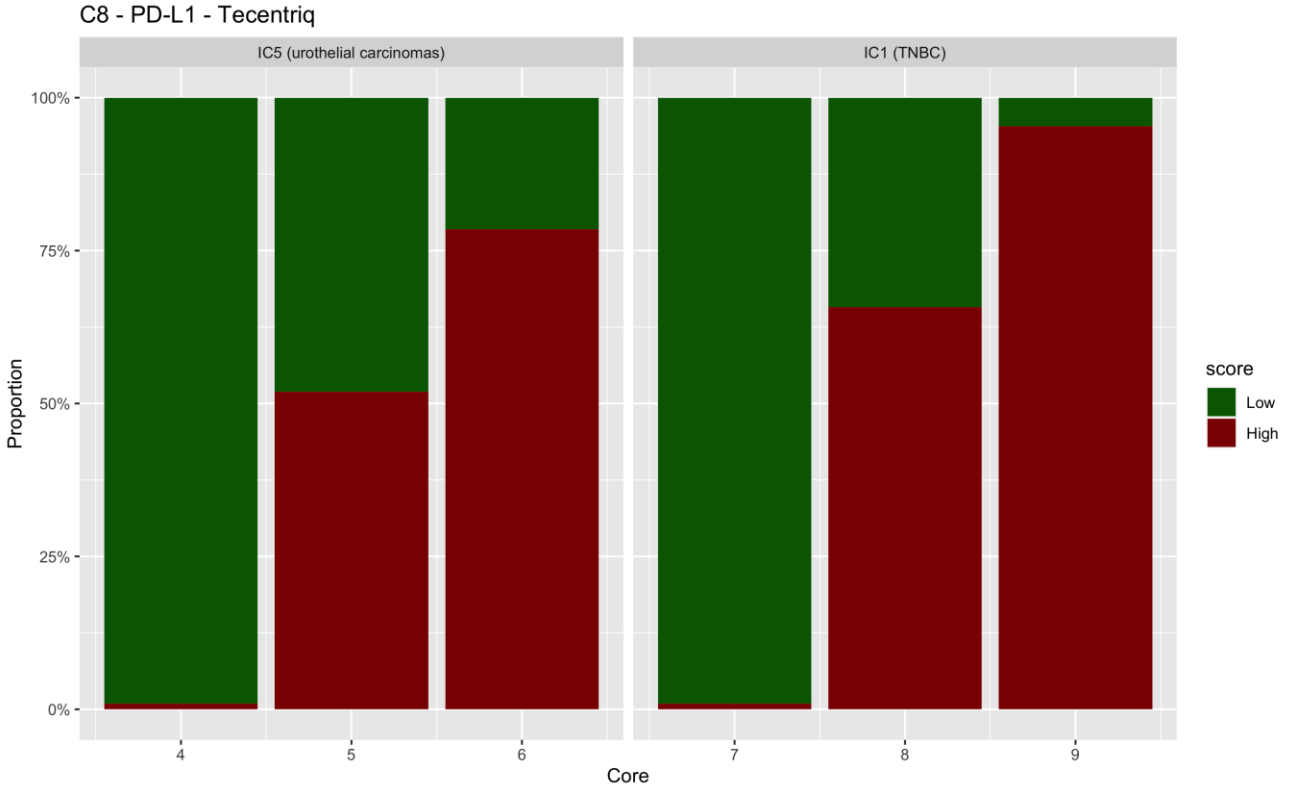
The Ventana PD-L1 SP142 assays 741-4860/740-4859 with predictive claim for TECENTRIQ® were used by 38% of the participants and provided a pass rate of 83% (39 of 47) when applying protocol settings in compliance with the vendor recommendations. The assays are locked for central protocol settings and 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 two assays, some laboratories submitted protocols with reported modified settings indicating change in incubation 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 two assays product no. 741-4860/740-4859 in this report and all protocols thus compiled as used by vendor recommended protocol settings as shown in Tables 2 and 3.

The Ventana PD-L1 SP142 assay 790-4860 without any predictive claim and available as an analytical or generic PD-L1 assay was used by 35% of the participants. This assay is based on same recommended protocol settings as the CDx products 741-4860/740-4859, 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 an increased 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. No obvious cause for this was found during the protocol analysis. If modifying the protocol, a similar pass rate was obtained compared to the CDx product 741-4860.

"Non-SP142" companion diagnostic assays as SP263 (Ventana), 22C3 pharmDx (Dako) laboratory developed (LD) tests based on either concentrated primary Abs or Ready-To-Use formats gave an overall significantly inferior performance and reduced pass rate compared to the SP142 assays from Ventana. In this assessment none with a sufficient staining result. The vast majority of the insufficient results were characterized by an extensive staining reaction of tumour cells compromising the scoring of PD-L1 expression in immune cells.
Similar observations were seen in both run C6 and C7, and these data indicate a challenge for the interchangeability of the Ventana SP142 assays with other PD-L1 companion diagnostic assays and LD assays designed and developed to primarily provide a staining pattern as characterized by e.g. the Dako 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 assay 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 pattern largely comparable to the Ventana 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 indicate 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 TECENTRIQ®, 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 addresses 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 1.



Graph 1. NordiQC PD-L1 run C8: Interpretation of IC in four urothelial carcinomas and four TNBC.

As seen in Graph 1, relatively high consensus rates were observed in core 4,6,7 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, tonsil was found to be a recommendable positive and negative tissue control and superior to placenta.

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. 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. Placenta did not contain the same dynamic PD-L1 expression range as seen in tonsil as throphoblasts typically showed a strong staining intensity due to high-level PD-L1 expression.

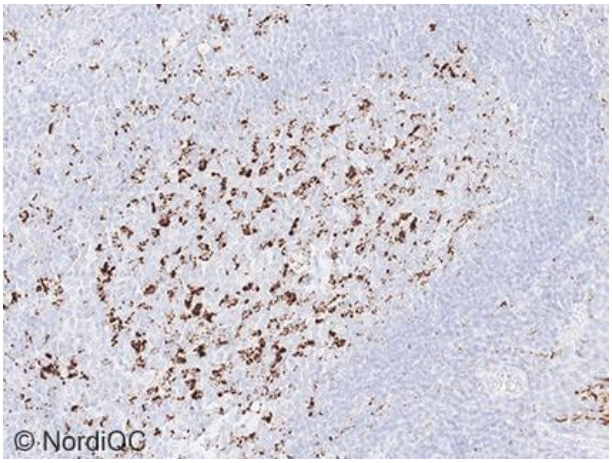


Fig. 1a
Optimal staining result of tonsil using the PD-L1 assay 741-4860 from Ventana, based on the rmAb clone SP142 following the recommended protocol settings. Same protocol used in Figs. 2a-6a.
Most germinal centre lymphocytes/macrophages and scattered interfollicular immune cells show a moderate to strong staining reaction.

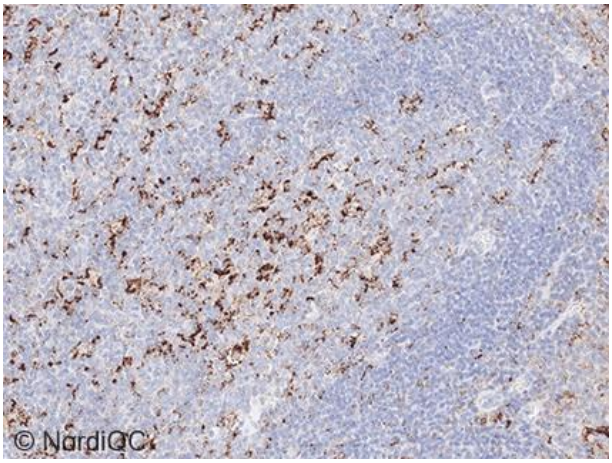


Fig. 1b
Staining result of tonsil using the mAb clone 22C3, diluted 1:40. The protocol was based on HIER in CC1 for 48 min., 60 min. incubation of the primary Ab and OptiView with Amplification as detection system. Same protocol in Figs. 2b-6b.
The intensity and proportion of immune cells is increased compared to the level obtained by the SP142 based assay. Compare with Fig. 1a – same area.

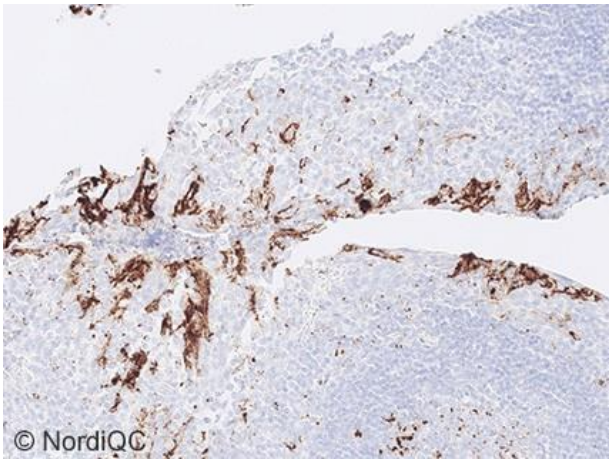


Fig. 2a
Optimal staining result of tonsil using same protocol as in Fig. 1a.
The crypt epithelial cells show an intense staining reaction, while superficial squamous epithelial cells being negative.

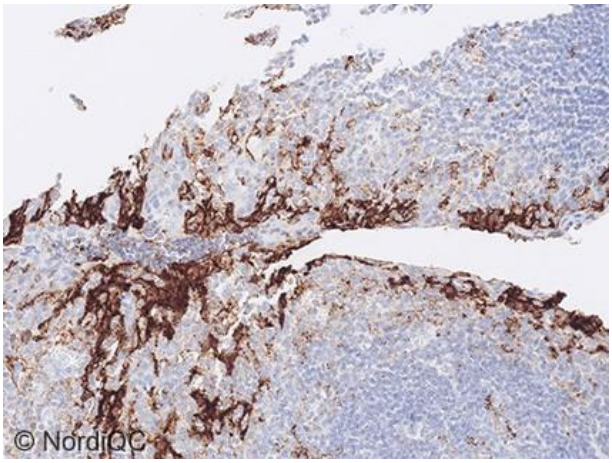


Fig. 2b
Staining result of tonsil using same protocol as in Fig. 1b.
The staining intensity of crypt epithelial cells is increased compared to the level obtained by the SP142 based assay. Compare with Fig. 2a – same area.

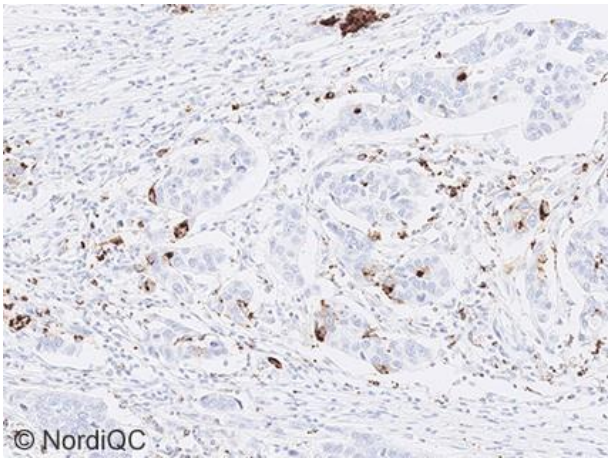


Fig. 3a
Optimal staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1a-2a. Virtually all tumour cells are negative and immune cells show a weak to moderate staining reaction giving an IC score of ≥5%.
The absence of staining reaction in the tumour cells facilitates the interpretation.

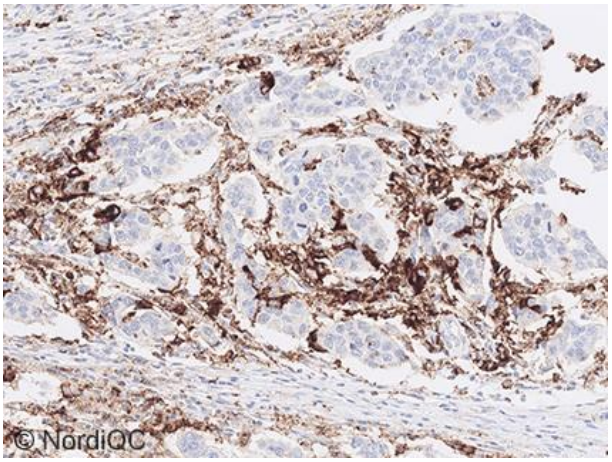


Fig. 3b
Staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1b-2b. Scattered tumour cells display a weak, granular membranous staining reaction and immune cells show a moderate to strong staining reaction. The intensity and proportion of immune cells is increased compared to the optimal result shown in Fig. 3a – same area. However, the PD-L1 status based on IC score is still concordant to reference standard level (≥5%).

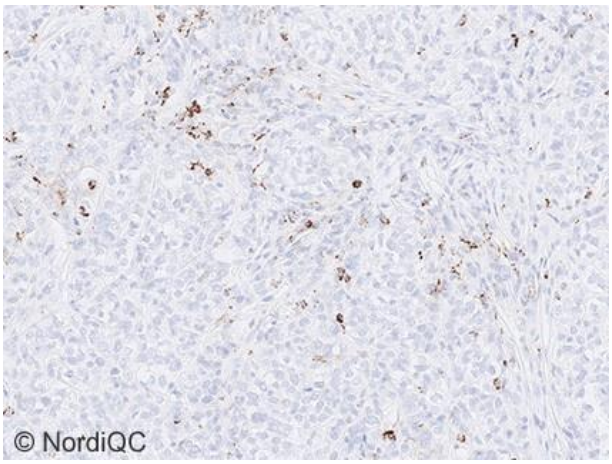


Fig. 4a
Optimal staining result of the TNBC, tissue core no. 9, using same protocol as in Figs. 1a-3a. A weak staining reaction is seen in scattered tumour cells. Immune cells display a moderate to strong staining reaction giving an IC score ≥1%.

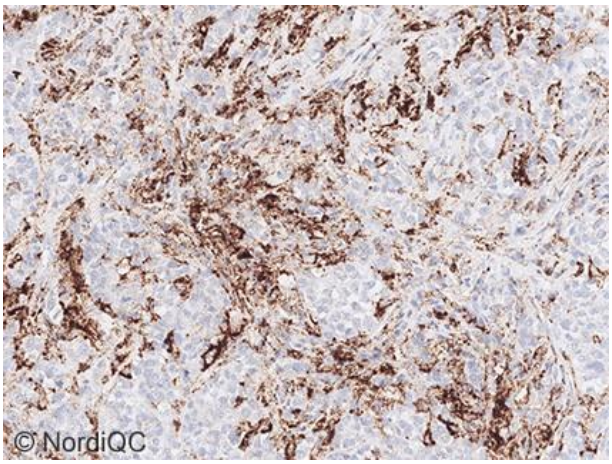


Fig. 4b
Insufficient staining result of the TNBC, tissue core no. 9, using same protocol as in Figs. 1b-3b. An excessive membranous staining reaction is seen in tumour cells compromising the identification and interpretation of PD-L1 reaction in the immune cells. Compare with optimal result shown in Fig. 4a – same area.

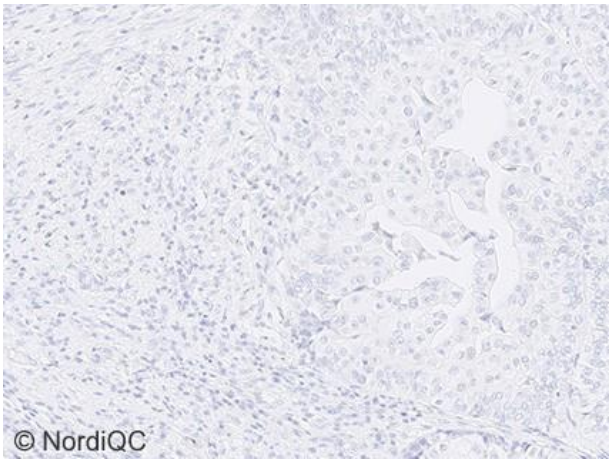


Fig. 5a
Optimal staining result of the urothelial carcinoma, tissue core no. 4, using same protocol as in Figs. 1a-4a. All immune cells are negative.

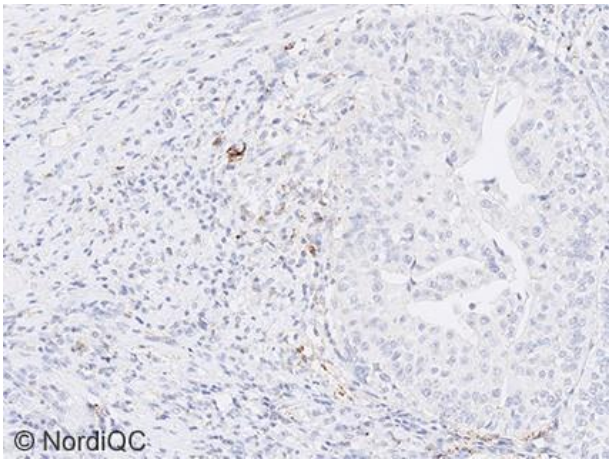


Fig. 5b
Insufficient staining result of the urothelial carcinoma, tissue core no. 4, using same protocol as in Figs. 1b-4b.
A weak but distinct staining reaction is seen in the immune cells expected to be negative. The IC score is close to 5%.
Compare with optimal staining result in Fig. 5a.

HLK/RR/LE/SN 05.01.2021