Assessment Run C7 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 eight 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, being needed.

Material

Table 1. **Content of the TMA used for the NordiQC PD-L1 TECENTRIQ® C7 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%
6. Urothelial carcinoma	≥5% (IC 5-10%)
7. Urothelial carcinoma	≥5% (IC 5-20%)
8. TNBC**	<1%
9. TNBC	<1%
10. TNBC	≥1% (IC 15-20%)***
11. 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.
*** Due to heterogeneity the IC category changed throughout one block to 1-5%.

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 as scoring 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 twentieth slide were 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) IHC assay (741-4860 Ventana).

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 8 carcinomas.</u>

Criteria for assessing a staining as Good 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 is still in the right category. The protocol may be optimized to ensure analytical accuracy and/or improved counter staining, morphology and signal-to-noise ratio.
IC score is still concordant to the NordiQC reference data in all 8 carcinomas.

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 false positive staining reaction of one of the included tissues. The protocol should be optimized.
IC score is **not** concordant to the NordiQC reference data in 1 of the 8 carcinomas.

Criteria for assessing a staining as Poor include:
The staining is considered very insufficient, e.g. because of a false negative or false positive staining reaction of more than one of the included tissues. Optimization of the protocol is urgently needed.
IC score is **not** found concordant to the NordiQC reference data in more than 1 of the 8 carcinomas.

A staining can also be assessed as **borderline/poor** in case the interpretation and scoring being significantly hampered by impaired morphology or excessive background reaction etc.

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 eight 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 C7	117
Number of laboratories returning PD-L1 TECENTRIQ® IHC	102 (87%)
Number of laboratories returning PD-L1 TECENTRIQ® scoring sheet	92 (90%)

The number of laboratories returning slides decreased in this run C7 compared to previous assessments, due to the COVID-19 pandemic. All slides returned after the assessment will be assessed, and receive advice if the result is insufficient, but will not be included in this report.

Results: 102 laboratories participated in this assessment and 54% 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 second NordiQC assessment of PD-L1 for **TECENTRIQ®**. The overall pass rate decreased significantly compared to the result obtained in run C6, 2019 (see Table 3).

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

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

Conclusion
This was the second NordiQC assessment of PD-L1 for **TECENTRIQ®** in urothelial carcinoma and TNBC in the companion module. 102 laboratories participated and a relatively low pass rate of 55% was observed being inferior to the result observed in run C6.
The PD-L1 SP142 companion diagnostic (CDx) IHC assay product no. 741-4860 from Ventana performed in compliance with vendor protocol recommendations was the most successful assay 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 93%. Other PD-L1 CDx assays as SP263 (741-4905, Ventana) and 22C3 (SK006/GE006) Dako being very successful in the NordiQC PD-L1 KEYTRUDA® assessments provided significantly lower pass rates of 29% and 20%, respectively. In total, it was observed that "non-SP142" based assays gave a pass rate of 15%. 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 C7, 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]**	29	Ventana/Roche	21	6	2	-	93%	72%
rmAb clone SP142, **741-4860 (LMPS)[4]**	7	Ventana/Roche	2	3	-	2	71%	29%
rmAb clone SP263, **741-4905 (VRPS)[3]**	6	Ventana/Roche	-	1	5	-	17%	-
rmAb clone SP263, **741-4905 (LMPS)[4]**	1	Ventana/Roche	-	1	-	-	-	-
mAb clone 22C3 pharmDX, **SK006 (VRPS)[3]**	2	Dako/Agilent	-	1	1	-	-	-
mAb clone 22C3 pharmDX, **SK006 (LMPS)[4]**	2	Dako/Agilent	-	-	2	-	-	-
mAb clone 22C3 pharmDX, **GE006 (LMPS)[4]**	1	Dako/Agilent	-	-	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	-	1	5	1	14%	-
mAb clone **E1L3N**	3	Cell Signaling	-	-	3	-	-	-
rmAb **CAL10**	3 1	Zytomed Biocare	-	1	3	-	-	-
rmAb clone **ZR3**	1	Gene Tech	-	-	1		-	-
Ready-To-Use antibodies[6]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP142, **790-4860 (VRPS)[3]**	15	Ventana/Roche	4	6	3	2	67%	27%
rmAb clone SP142, **790-4860 (LMPS)[4]**	11	Ventana/Roche	4	3	3	1	64%	36%
rmAb clone SP263, **790-4905 (VRPS)[3]**	5	Ventana/Roche	-	-	5		0%	-
rmAb clone SP263, **790-4905 (LMPS)[4]**	2	Ventana/Roche	-	2	-		-	-
rmAb clone 73-10, **PA0832 (VRPS)[3]**	2	Leica Biosystems	-	-	2	-	-	-
rmAb clone MXR003, **RMA-0732**	4	Maixin	-	-	4	-	-	-
Total	102		31	25	40	6		
Proportion			30%	25%	39%	6%	55%	

1) Proportion of sufficient stains (optimal or good) (≥5 asessed protocols).
2) Proportion of optimal results (≥5 asessed 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, 23 of 36 (67%) protocols were assessed as optimal. Protocols with optimal results were based on Heat Induced Epitope Retrieval (HIER) in Cell Conditioning 1 (CC1) (efficient heating time 16-70 min.), 16 or 24 min. incubation of the primary Ab and OptiView with OptiView Amplification as detection system and performed on the BenchMark GX/Ultra. Using these protocol settings, 27 of 29 (93%) laboratories 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**	27/29 (93%)	21/29 (72%)	5/7 (71%)	2/7 (29%)
Ventana BenchMark XT, GX, Ultra rmAb SP263, **741-4905**	1/6 (17%	0/6 (0%)	1/1	0/1
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	1/2	0/2	0/2	0/2
Dako Omnis mAb 22C3 pharmDX, **GE006**	-	-	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, 8 of 26 (31%) protocols were assessed as optimal. Protocols with optimal results were typically based on HIER in CC1 (efficient heating time 36-64 min., 16-40 min. incubation of primary Ab and OptiView with or without OptiView Amplification as detection system. Using these settings, 15 of 21 (71%) produced a sufficient staining result.

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

In this second NordiQC run for PD-L1 TECENTRIQ® in the companion module C7, a pass rate of 55% 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 significantly reduced compared to the result obtained in run C6. The central parameters potentially affecting pass rates in IHC proficiency schemes were identical in the two runs. Of critical importance in both run C6 and C7, 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 C7 most likely was more challenging compared to run C6, especially for PD-L1 assays less selective for the identification of 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 52% (24 of 46) of the insufficient staining results. In 26% (12 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. The remaining 22% (10 of 46) of the insufficient results were caused by a false positive staining result. Table 4 shows the main characteristics of insufficient results in the two NordiQC PD-L1 TECENTRIQ® runs.

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

	False Negative*	**False Positive****	**"Technical"*****
C6 (24% insufficient)	20%	-	80%
C7 (45% insufficient)	26%	22%	52%
Average (35% insufficient)	**23%**	**11%**	**66%**

* 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.

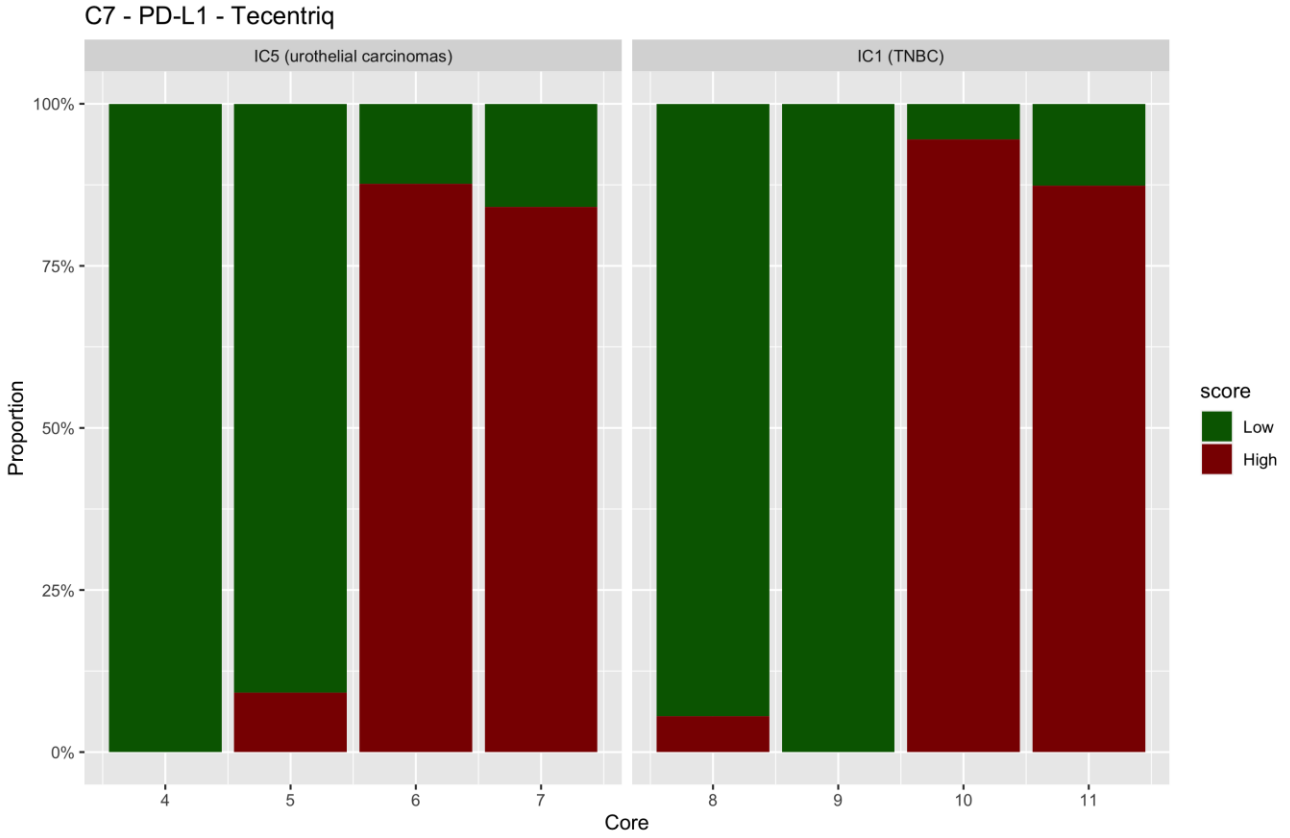
The Ventana PD-L1 SP142 assay 741-4860 with predictive claim for TECENTRIQ® was used by 35% of the participants and provided a pass rate of 93% (27 of 29) when applying protocol settings in compliance with the vendor recommendations. The assay is based on a "locked" protocol based on HIER in CC1 for 48

min., incubation in primary Ab for 16 min. (Ultra/XT) or 24 min. (GX) and use of OptiView with Amplification as detection system. Despite the locked protocol conditions some laboratories submitted protocols with reported modified settings indicating change in incubation time of primary Ab and other detection system applied – e.g. UltraView and OptiView without Amplification. Protocols based on these changes were listed as LMPS in Table 2 and provided a reduced pass rate and proportion of optimal results compared to the Ventana VRPS. HIER time in CC1 was not included in the separation of VRPS versus LMPS, as many laboratories faced difficulties to extract the data for this.

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 26% of the participants. This assay is based on same recommended protocol settings as for the CDx product 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 790-4980 format gave a reduced pass rate and proportion of optimal results compared to the corresponding CDx format of the same clone as seen in Table 2. This was observed for both VRPS and LMPS. The feature of insufficient results was typically characterized by a completely false negative reaction in one or all tissue cores. No plausible explanations correlated to the protocols applied could explain the aberrant result, and most likely caused by "technical system issues" related to e.g. the instruments, reagents or other general components of the IHC test.

"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 total for this group, a pass rate of 18% (7 of 40) was observed, none optimal. 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. In addition, some insufficient results were also caused by a false positive result in the urothelial carcinoma, tissue core no. 5, expected to be PD-L1 negative as defined by the NordiQC reference standard method based on the SP142 CDx assay for TECENTRIQ®. Similar observations were seen in run C6, 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.

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 C7: Interpretation of IC in four urothelial carcinomas and four TNBC.
As seen in Graph 1, relatively high consensus rates were observed. 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 an insufficient technical mark. In the technical assessment it was observed that some protocols provided a false positive staining reaction in the urothelial carcinoma, tissue core no. 5. This was also reflected in the scoring results from the participants, where more than 5% classified tissue core as PD-L1 IHC IC ≥5%.

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 only throphoblasts typically showed a strong staining intensity due to high-level PD-L1 expression.
However placenta might be a beneficial supplemental positive tissue control to monitor the performance of the SP142 assay based on OptiView with Amplification kit. Both in this run and C6, a reduced and insufficient level of analytical sensitivity of the SP142 assay typically only provided a patchy focal reaction of throphoblasts and not the expected diffuse moderate to strong reaction in all cells. The observation has to be further investigated in order to confirm or dispute the observations.

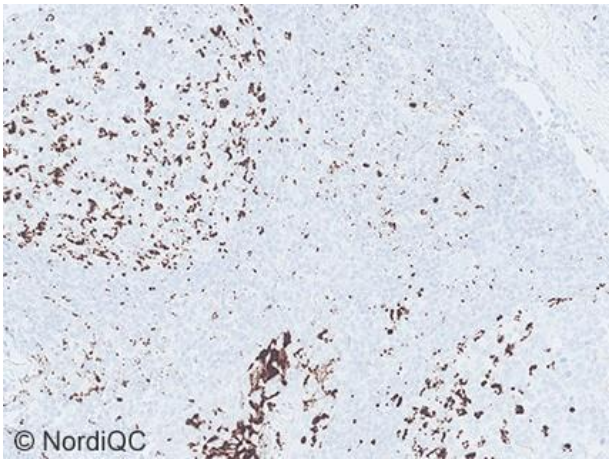


Fig. 1a
Optimal staining result of tonsil using the PD-L1 assay 741-4860, Ventana, based on the rmAb clone SP142 following the recommended protocol settings. Same protocol used in Figs. 2a-6a.
Most germinal centre lymphocytes/macrophages and scattered interfollicular immune cells show a moderate to strong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

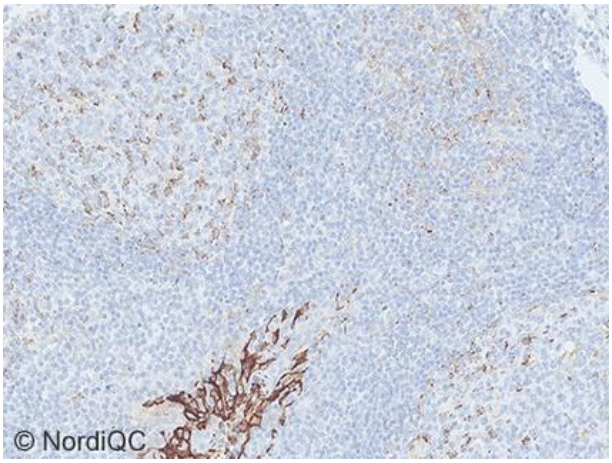


Fig. 1b
Insufficient staining result of tonsil using the PD-L1 assay, 741-4905, Ventana based on the rmAb clone SP263 following the recommended protocol settings. Same protocol in Figs. 2b-4b.
The intensity and proportion of immune cells is reduced compared to the level obtained by the SP142 based assay. Compare with the optimal result shown in Fig. 1a – same area.

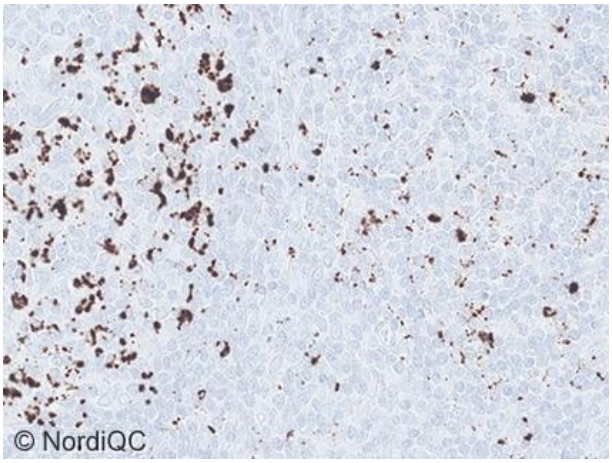


Fig. 2a
Optimal staining result of tonsil (x400) using same protocol as in Fig. 1a.
The immune cells demonstrated show a moderate to strong coarse granular staining reaction.

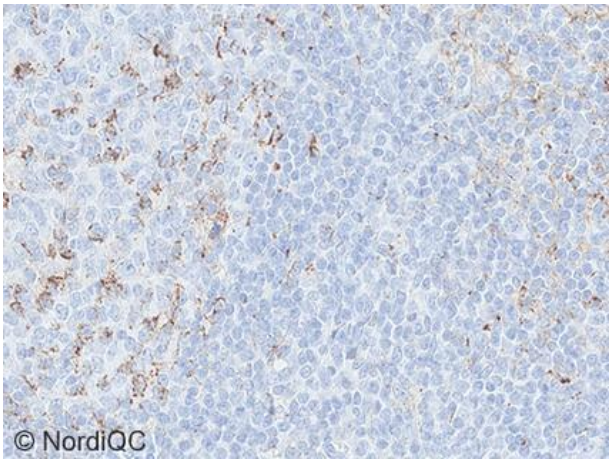


Fig. 2b
Insufficient staining result of tonsil (x400) using same protocol as in Fig. 1b.
The staining intensity of especially the interfollicular immune cells is reduced compared to the level obtained by the SP142 based assay. Compare with optimal result shown in Fig. 2a – same area.

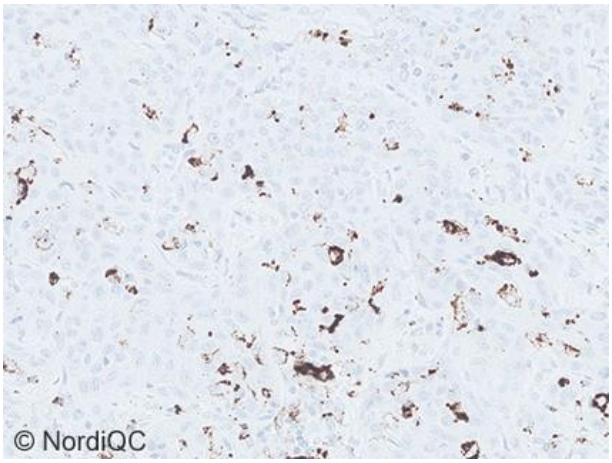


Fig. 3a
Optimal staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1a-2a. The tumour cells are negative and immune cells show a moderate to strong staining reaction giving an IC score ≥5%.

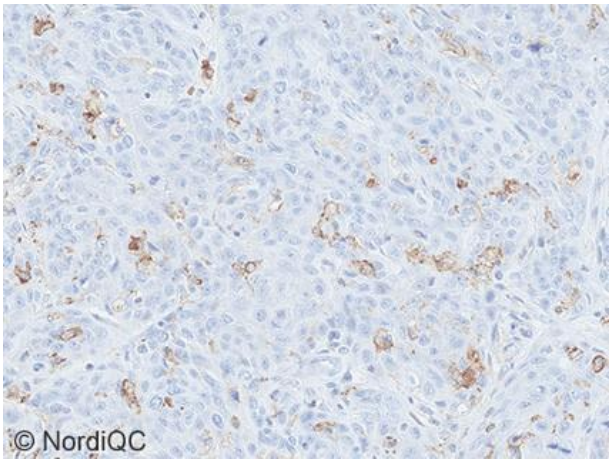


Fig. 3b.
Staining result of the urothelial carcinoma, tissue core no. 6, using same protocol as in Figs. 1b-2b. The tumour cells are negative and immune cells show a predominantly weak to moderate staining reaction. The intensity and proportion of immune cells is reduced compared to the optimal result shown in Fig. 3a but giving an IC score ≥5%. However, also compare with the result shown in Fig. 4b.

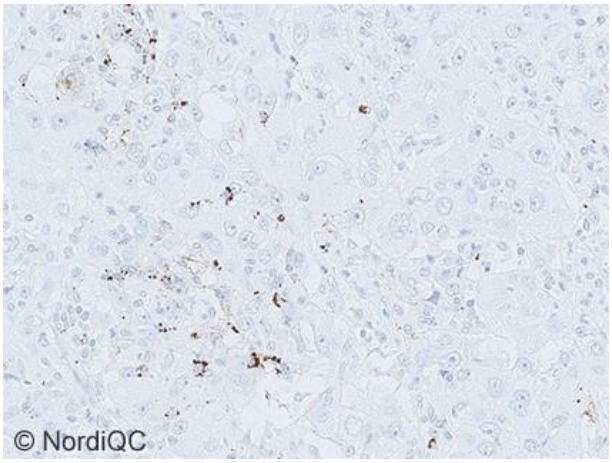


Fig. 4a
Optimal staining result of the TNBC, tissue core no. 11, using same protocol as in Figs. 1a-3a. Virtually all tumour cells are negative and immune cells show a weak to strong granular staining reaction giving an IC score ≥1%.
The absence of staining reaction in the tumour cells facilitates the interpretation.

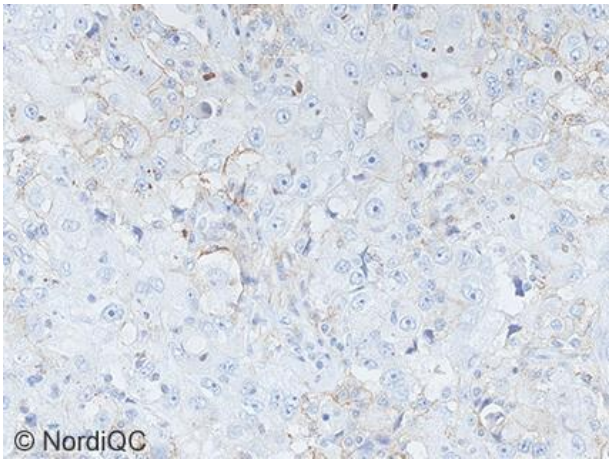


Fig. 4b
Insufficient staining result of the TNBC, carcinoma tissue core no. 11, using same protocol as in Figs. 1b-3b. A distinct staining reaction is obtained in the majority of tumour cells hampering the interpretation of PD-L1 in the immune cells being faintly demonstrated.
Compare with the optimal result shown in Fig. 4a – same area.

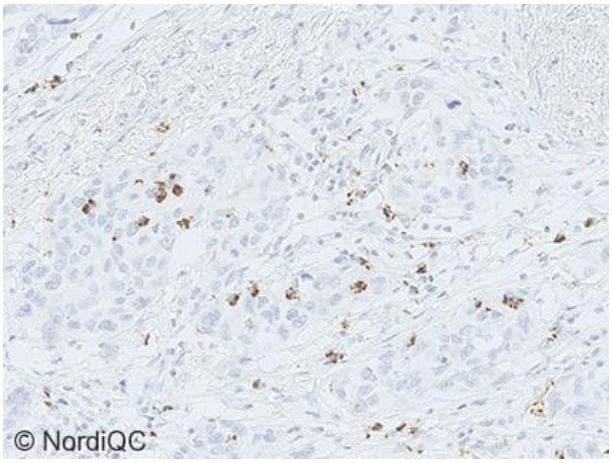


Fig. 5a
Optimal staining result of the urothelial carcinoma, tissue core no. 7, using same protocol as in Figs. 1a-4a. A faint staining reaction is seen in scattered tumour cells. Immune cells display a moderate to strong staining reaction giving an IC score ≥5%.

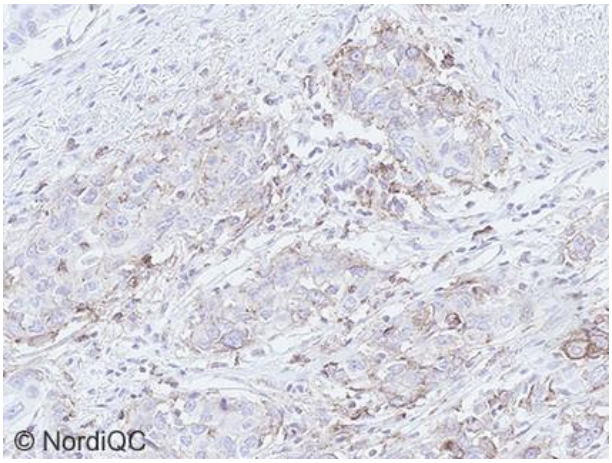


Fig. 5b
Insufficient staining result of the urothelial carcinoma, tissue core no. 7, using the Maixin RTU format RMA-0732 of rmAb clone MXR003. An excessive membranous staining reaction is seen in tumour cells compromising the interpretation of immune cells. Compare with optimal result shown in Fig. 5a – same area.

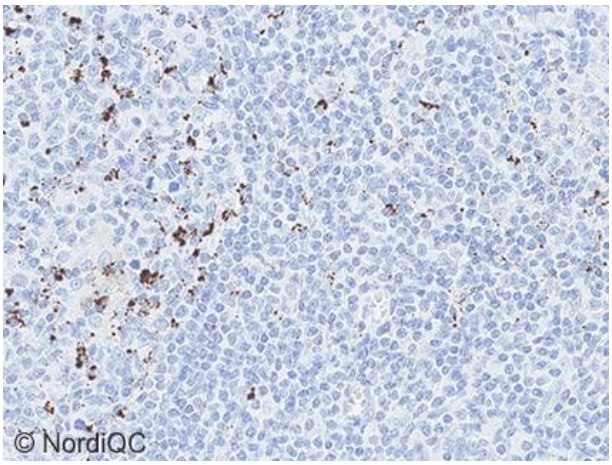


Fig. 6a
Insufficient staining result of tonsil, using the PD-L1 assay 741-4860, Ventana, based on the rmAb clone SP142 following the recommended protocol settings.
The proportion and intensity of immune cells is reduced compared to the result expected and shown in Fig. 1a using same protocol. The discrepancy of the staining outcome most likely caused by technical issues e.g. related to the IHC instrument or malfunctioning general IHC components as e.g. precipitations in prep kits etc.

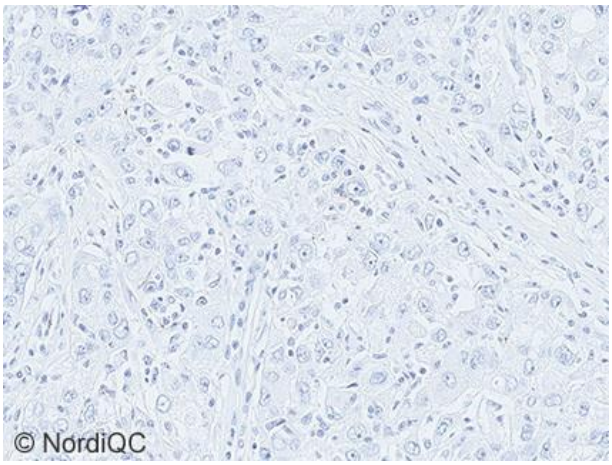


Fig. 6b
Insufficient staining result of the TNBC, tissue core no. 11, using same protocol as in Fig. 6a. Only a faint staining reaction is seen single immune cells and the IC score is <1% and thus false negative.
Compared to optimal staining result in Fig. 4a.

SN/LE/HLK/RR 07.07.2020