Assessment Run C8 2020
PD-L1 KEYTRUDA®

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

Material
Table 1. **Content of the TMA used for the NordiQC PD-L1 KEYTRUDA® C8 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% / High; 50-70%
6. NSCLC	TPS: High; 90-100%
7. Urothelial carcinoma	CPS: <10
8. Urothelial carcinoma	CPS: ≥10; 20-80 IC***
9. Urothelial carcinoma	CPS: ≥10; 40-100 IC+TC***

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

All tissues were fixed in 10% neutral buffered formalin.

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

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

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

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

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

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

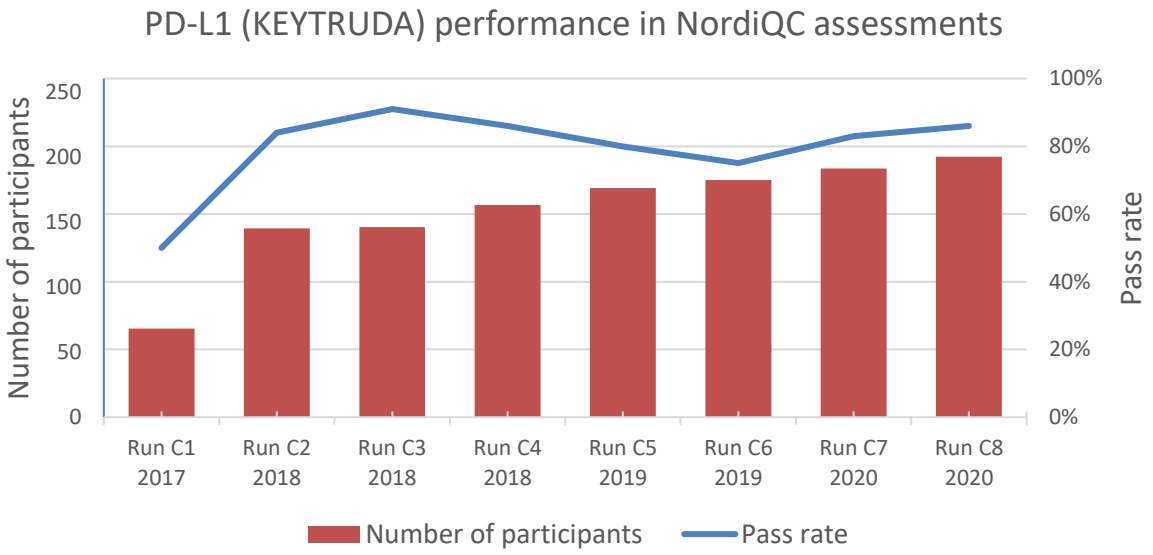
An IHC result can also be assessed as **borderline/poor** related to technical artefacts, e.g. poor signal-to-noise ratio, excessive counterstaining, impaired morphology and/or excessive staining compromising the scoring.

Participation

Number of laboratories registered for PD-L1 KEYTRUDA IHC C8	212
Number of laboratories returning PD-L1 KEYTRUDA IHC	200 (94%)
Number of laboratories returning PD-L1 scoring sheet	183

Results: 200 laboratories participated in this assessment and 86% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Fig. 1.

Fig. 1. **Proportion of sufficient results for PD-L1 (lung)/KEYTRUDA in the eight NordiQC runs performed**



Performance history
This was the eight NordiQC assessment of PD-L1 TPS/CPS KEYTRUDA®. A slightly improved pass rate has been obtained in the last three assessments. The number of new participants seems to be consistently increasing.

Conclusion
In this run and similar to observations seen in previous runs in the PD-L1 companion module, 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. This resulted in a too low TPS/CPS changing PD-L1 status in one or more of the carcinomas included. Several companion diagnostic (CDx) assays and laboratory developed (LD) assays based on concentrated Abs as mAb clone 22C3 or Ready-To-Use formats of the rmAb clones 73-10 and MAB-0854 could all provide optimal results.
The PD-L1 IHC CDx assays, 22C3 GE006 from Dako/Agilent and SP263 741-4905 / 740-4907 from Ventana/Roche performed in concordance to the product guidelines, were most successful providing a high proportion of sufficient and optimal results being superior to both the CDx assay 22C3, SK006 Dako/Agilent and LD assays based on concentrated Abs.

Table 2. **Assessment marks for IHC assays and antibodies run C8, PD-L1 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]**	47	Ventana/Roche	37	9	1	-	98%	79%
rmAb clone SP263, **740-4907 (VRPS)[3]**	5	Ventana/Roche	5	-	-	-	100%	100%
rmAb clone SP263, **740-4907 (LMPS)[4]**	1	Ventana/Roche	1	-	-	-	-	-
rmAb clone SP142, **740-4859/741-4860 (VRPS)[3]**	2	Ventana/Roche	-	-	1	1	-	-
mAb clone 22C3 pharmDX, **SK006 (VRPS)[3]**	18	Dako/Agilent	7	6	4	1	72%	39%
mAb clone 22C3 pharmDX, **SK006 (LMPS)[4]**	16	Dako/Agilent	3	11	2	-	88%	19%
mAb clone 22C3 pharmDX, **GE006 (VRPS)[3]**	17	Dako/Agilent	12	5	-	-	100%	71%
mAb clone 22C3 pharmDX, **GE006 (LMPS)[4]**	11	Dako/Agilent	2	7	2	-	81%	18%
rmAb clone 28-8 pharmDX, **SK005 (VRPS)[3]**	3	Dako/Agilent	1	2	-	-	-	-
Antibodies[5] for laboratory developed PD-L1 assays, concentrated antibodies	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
mAb clone **22C3**	40	Dako/Agilent	8	21	8	3	73%	20%
mAb clone **E1L3N**	4	Cell Signaling	-	3	1	-	-	-
rmAb clone **BP6141**	1	Bailing Biotechnology	-	1	-	-	-	-
rmAb **CAL10**	4 1	Biocare Zytomed Systems	2	-	1	2	20%	20%
rmAb clone **QR1**	1	Biocyc	-	1	-	-	-	-
rmAb clone **SP142**	1	Spring Biosystems	1	-	-	-	-	-
rmAb clone **ZR3**	1 1	Gene Tech Zeta Corporation	1	1	-	-	-	-
Ready-To-Use antibodies[6]	n	Vendor	Optimal	Good	Borderline	Poor	Suff.[1]	OR[2]
rmAb clone SP263, **790-4905[6] (VRPS)[3]**	13	Ventana/Roche	11	2	-	-	100%	85%
rmAb clone SP263, **790-4905[6] (LMPS)[4]**	9	Ventana/Roche	5	3	1	-	89%	56%
rmAb clone 73-10, **PA0832 (VRPS)[3]**	1	Leica Biosystems	1	-	-	-	-	-
rmAb clone MX070C, **MAB-0854**	3	Maixin	2	1	-	-	-	-
Total	200		99	73	21	7		
Proportion			49%	37%	11%	3%	86%	

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, 37 of 47 (79%) 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 64 min., 16 min. incubation of primary Ab and OptiView as detection system. Using these protocols settings, 46 of 47 (98%) laboratories produced a sufficient staining result (optimal or good).

SP263 (740-4907, Ventana/Roche): In total, 6 of 6 (100%) protocols were assessed as optimal. This product has a locked protocol on all BenchMark platforms and cannot be changed. The protocol is based on HIER in CC1 for 64 min., 16 min. incubation of primary Ab and OptiView as detection system. 5 laboratories used this on BenchMark Ultra and GX, all with optimal results. 1 used the product off-label on a Bond (Leica Biosystems), also producing a result assessed as optimal.

PD-L1 IHC 22C3 pharmDx (SK006, Dako/Agilent): In total, 10 of 34 (29%) 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, 13 of 18 (72%) laboratories produced a sufficient staining result.
The SK006 was frequently used by modified protocol settings e.g. mitigation to other platform as Dako Omnis, Ventana BenchMark or performed manually.

PD-L1 IHC 22C3 pharmDx (GE006, Dako/Agilent): In total, 14 of 28 (50%) protocols were assessed as optimal. Protocols with optimal results were 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, 17 of 17 (100%) laboratories produced a sufficient staining result.

PD-L1 IHC 28-8 pharmDx (SK005, Dako/Agilent): In total, 1 of 3 protocols was assessed as optimal. The protocol with optimal result was based on HIER using EnVision™ Flex TRS low pH 6.1 at 97°C for 20 min. (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, 2 of 3 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**	46/47 (98%)	37/47 (79%)	-	-
Ventana BenchMark Ultra rmAb SP263, **740-4907**	5/5 (100%)	5/5 (100%)	-	-
Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006**	13/18 (72%)	7/18 (39%)	6/7 86%	1/7 14%
Dako Omnis mAb 22C3 pharmDX, **GE006**	17/17 (100%)	12/17 (71%)	5/5 (100%)	2/5 (20%)
Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005**	3/3	1/3	-	-

*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**: 8 of 40 (20%) protocols were assessed as optimal of which six were stained on the BenchMark (Ventana) and two on the Omnis (Dako) platform.
On BenchMark XT/Ultra (Ventana), the protocols providing optimal results were based on a titre of 1:20-50, primary Ab incubation time of 60-120 min., HIER in CC1 (efficient heating time 48-80 min.) and OptiView as detection system. Using these protocol settings, 12 of 12 (100%) laboratories produced a sufficient staining result.
On Omnis (Dako), the protocols providing optimal results were typically based on a titre of 1:20 of the primary Ab, incubation time of 40 min., HIER in TRS low pH 6.1 (Dako) at 97°C (efficient heating time 40 min.) and EnVision FLEX+ as detection system. Using these protocol settings, 4 of 4 (100%) laboratories produced a sufficient staining result.

rmAb **CAL10**: 2 of 5 protocols (40%) were assessed as optimal. One protocol was based on HIER using an alkaline buffer (BERS2, Leica) at 99°C for 20 min. The rmAb clone CAL10 was diluted 1:50, incubated for 20 min. at room temp. and visualized using a 3-layer detection system (Leica Refine) and performed on a Leica Bond III platform.
The other protocol was based on HIER in Diva Declaoker in a pressure cooker at 110°C for 15 min. The rmAb clone CAL10 was diluted 1:200 for 30 min. at room temp. and visualized by MACH4 (Biocare) and performed on a Biocare IntelliPATH platform.

rmAb **ZR3**: One protocol was assessed as optimal. The protocol was based on HIER using Tris-EDTA/EGTA pH 9 at 95°C for 15 min. in a waterbath. The rmAb clone ZR3 was diluted 1:150, incubated for 50 min. at room temp. and visualized using a 2-layer detection system (GT Vision, GeneTech) and performed on a GeneTech Genestainer.

rmAb **SP142**: One protocol provided an optimal result. The protocol was based on HIER using a TRS low pH 6.1 buffer (Dako) at 100°C for 20 min. in PT link. The rmAb clone SP142 was diluted 1:400, incubated for 30 min. at room temp. and visualized using a 3-layer detection system (Thermo, Quanto) and performed on a Dako Autostainer Link 48 platform.

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

Concentrated antibodies	**Ventana/Roche BenchMark GX/XT/Ultra**		**Dako/Agilent Autostainer**		**Dako/Agilent Omnis**		**Leica 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**	6/12 (50%)	-	-	0/2	0/2	2/4	0/5	-

*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, 16 of 22 (73%) protocols provided an optimal result. Protocols with optimal results were typically based on HIER in CC1, efficient heating time 52-64 min., 16-20 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra or XT. Using these protocols settings, 19 of 19 (100%) laboratories produced a sufficient staining result (optimal or good).

rmAb **73-10** (PA0832, Leica): One protocol was assessed as optimal. The protocol was based on HIER in Bond Epitope retrieval Solution 1 for 20 min., 15 min. incubation of the primary Ab, Refine as detection system and performed on Bond III.

rmAb **MX070C** (MAB-0854, Maixin): 2 out of 3 protocols provided an optimal result (67%). The protocols were based on HIER in Tris-EDTA/EGTA pH 9 for 20-30 min., 30 min. incubation of the primay Ab, DAB Titan or Titan Super as detection system and performed on Lumatas Autostaining System. Using these protocol settings 3 of 3 (100%) 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 3 NSCLCs, 3 urothelial carcinomas, 2 tonsils and 1 placenta. The NSCLCs were selected to comprise 1 NSCLC for each TPS category: TPS negative (<1% PD-L1 positive tumour cells), TPS low (≥1-49%) and TPS high (≥50%). The urothelial carcinomas were selected to comprise 1 carcinoma with CPS<10 and 2 carcinomas with CPS≥10 - one with PD-L1 expression primarily in immune cells and one with PD-L1 expression in both tumour cells and immune cells, respectively. Reference slides throughout the individual TMA blocks (interval at each twenty-fifth slide) were stained using the companion diagnostic assays 22C3 pharmDX SK006 (Dako/Agilent), 22C3 pharmDx GE006 (Dako/Agilent) and SP263 741-4905 (Ventana/Roche). 22C3 pharmDX SK006 (Dako/Agilent) was used to characterize PD-L1 for both TPS and CPS levels, whereas 22C3 pharmDx GE006 and SP263 for were only used to characterize TPS (reflecting the EU/FDA approved predictive claims for KEYTRUDA® at the assessment). In total, eight identical TMA blocks were constructed and used for this assessment.
Reviewing the reference slides from the blocks, heterogenic expression of PD-L1 was seen in one of the tumor cores. In the NSCLC, tissue core no. 5, initially scored as TPS low (≥1-49%), slides with TPS high ≥50% was 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 eighth NordiQC assessment for PD-L1 for (lung)/KEYTRUDA®, the prevalent feature of an insufficient staining result was a false negative staining result, being observed in 75% (21 of 28) of the insufficient results. As shown in Table 5, a false negative staining result has been the most common reason for insufficient staining results in all NordiQC PD-L1 (lung)/KEYTRUDA® assessments with an average occurrence of 76%. In this run, 25% (7 of 21) of the insufficient results were caused by technical issues as poor-signal-to-noise ratio or excessive cytoplasmic staining reaction compromising the scoring of the PD-L1 status in one or more of the carcinomas. No false positive results were observed.

Table 5. **Characteristics of insufficient results in the eight NordiQC PD-L1 (lung)/KEYTRUDA® runs.**

	False Negative*	False Positive**	"Technical"***
C1 (50% insufficient)	68%	15%	17%
C2 (16% insufficient)	68%	23%	5%
C3 (9% insufficient)	100%	0%	0%
C4 (14% insufficient)	74%	4%	22%
C5 (20% insufficient)	59%	18%	23%
C6 (25% insufficient)	82%	11%	7%
C7 (17% insufficient)	78%	6%	16%
C8 (14% insufficient)	75%	0%	25%
Average (21% insufficient)	**76%**	**10%**	**14%**

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

In run C8 an increasing proportion of 37% of the participants obtained a score as good. In 52% of these (38 of 73), this was due to a general weak staining result or a reduced TPS/CPS, but with no change of TPS/CPS-category in any of the carcinomas and thus still an accurate PD-L1 status for treatment decision. In 10% (7 of 73) an increased TPS/CPS score was observed compared to the level expected, but again without any change in the PD-L1 status. 14% (10 of 73) of the results assessed as good were characterized by poor signal-to-noise ratio, impaired morphology or excessive counterstaining.
And finally, in 8% (6 of 73) a coarse granular staining reaction compromising the evaluation of the membranous staining reaction was observed. This pattern was only seen for protocols based on OptiView with amplification kit.

In this assessment and in concordance to the previous runs C6 and C7, virtually all insufficient results were related to incorrect TPS categories in one or more of the NSCLCs, whereas the CPS scores only were affected in a few cases of the urothelial carcinomas. PD-L1 IHC demonstration in the assessments with combined tumour material has thus been more successful in urothelial carcinomas versus NSCLCs. No plausible reasons for this difference can be identified. Expression levels in the materials used for the assessments in combination with different cut-off values and scoring methods might have favoured consistent PD-L1 demonstration in urothelial carcinomas. 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.

The NSCLC, tissue core no. 5, was most challenging to obtain an optimal result and required a carefully calibrated and reproducible protocol. Virtually all false negative results were seen in the NSCLC, tissue core no. 5, changing the TPS category compared to the level expected and defined by the NordiQC reference standard methods. In contrast, virtually all protocols provided the expected PD-L1 status in both the NSCLC, tissue core no. 6, characterized by NordiQC to show a strong membranous staining reaction in all tumour cells and in the two urothelial carcinomas with CPS≥10.

The Ventana/Roche PD-L1 IHC assays 741-4905 and 740-4907 for BenchMark (Ultra/XT/GX) with predictive claims, based on the SP263 clone, were used by 27% of the participants and provided a pass rate of 98% when applied by protocol settings in compliance with vendor recommendations. The assays are locked for central protocol settings and based on HIER in CC1 for 64 min., incubation in primary Ab for 16 min. and use of OptiView as detection system. Despite the locked protocol conditions for the two assays, some laboratories submitted protocols with reported modified settings typically indicating a change for HIER and/or incubation time of primary Ab. The different protocol settings submitted were disregarded for the two assays product no. 741-4905 and 740-4907 in this report and all protocols thus compiled as used by vendor recommended protocol settings as shown in Tables 2 and 3.

The Dako/Agilent 22C3 pharmDx assay SK006 for Autostainer Link 48 provided an overall pass rate of 79% (27 of 34). Applying the recommended protocol settings from Dako, the pass rate was 72% (13 of 18) and 39% optimal (7 of 18) (see Tables 2 and 3). The pass rate and proportion of optimal results was reduced compared to previous levels seen in the NordiQC assessments and also to both the corresponding Dako/Agilent 22C3 pharmDx assay for Omnis and the Ventana/Roche PD-L1 CDx assays based on SP263. The results assessed as "Good" were all characterized either by a generally weak staining reaction and/or a reduced number of cells being demonstrated compared to the level expected, but still no impact on PD-L1 status in any of the carcinomas. The insufficient results were caused by a significantly reduced proportion of cells demonstrated and a change of the PD-L1 status in one or more of the carcinomas. No plausible reason for the reduced analytical sensitivity and accuracy could be identified.

The Dako/Agilent 22C3 pharmDx assay GE006 for Omnis provided an overall pass rate of 92% (26 of 28). Applying the recommended protocol settings from Dako, the pass rate was 100% (17 of 17) and 71% optimal (12 of 17) (see Tables 2 and 3).
Similar to the data generated in runs C6 and C7, it was observed that the PD-L1 22C3 GE006 assay for Omnis was more successful compared to 22C3 pharmDx SK006. Cumulated data for the 3 successive runs has shown a pass rate of 100% (40 of 40) for laboratories using GE006 by vendor recommended protocol settings. In comparison a pass rate of 79% (41 of 52) for laboratories using SK006 has been obtained. The different pass rates observed have to be taken with caution due to relatively few data observations, but a clear trend so far has been observed in the three runs.
In this context it has to be emphasized that the 22C3 GE006 assay for Omnis is only validated by the vendor for PD-L1 status and predictive claim in NSCLC with TPS as scoring system and at present not validated for any indication with CPS as scoring system including urothelial carcinoma.

The Dako/Agilent pharmDx SK005 28-8 for Autostainer Link 48 was used by three laboratories. All used the Dako recommended protocol settings and two of three with a sufficient result. One result was assessed as optimal.

Grouped together, and using vendor recommended protocol settings, the CE IVD approved PD-L1 IHC assays with predictive claims - irrespective of indication and drug associated and e.g. both for KEYTRUDA® and other drug as e.g. OPDIVO, 28-8 SK005, Dako Agilent - provided a pass rate of 91% (84 of 92) and 67% being optimal (62 of 92). These levels indicate a possibility of interchangeability between the assays for PD-L1 status for KEYTRUDA® using the present cut-off values and scoring methods for TPS/CPS in the two indications addressed in this module. This must be validated by end-user according to local regulations.
The Ventana CDx assays based on SP142 was used by two participants and both results submitted assessed as insufficient. Several publications inclusive Blueprint studies 1 and 2 (Hirsch, Tsao et al) have indicated poor analytical concordance for SP142 compared to the other CDx assays for TPS.

Laboratory developed (LD) assays either based on a concentrated Ab, an RTU Ab without any predictive claim or a companion diagnostic assay not used strictly accordingly to vendor recommendations were applied by 54% (108 of 200) of the participants. For this group a pass rate of 81% (88 of 108) was observed, 34% optimal (37 of 108).
LD assays based on concentrated formats provided a relatively low pass rate of 72% (39 of 54), only 22% being optimal (12 of 54) (see Table 1).

Protocols based on RTU Abs without predictive claims as rmAb clones SP263 (790-4905, Ventana/Roche), 73-10 (PA0835, Leica) and MX070C (MAB-0854, Maixin) were more successful. For LD based assays on RTU Abs, the overall pass rate and proportion of sufficient and optimal results were 96% and 73%, respectively.

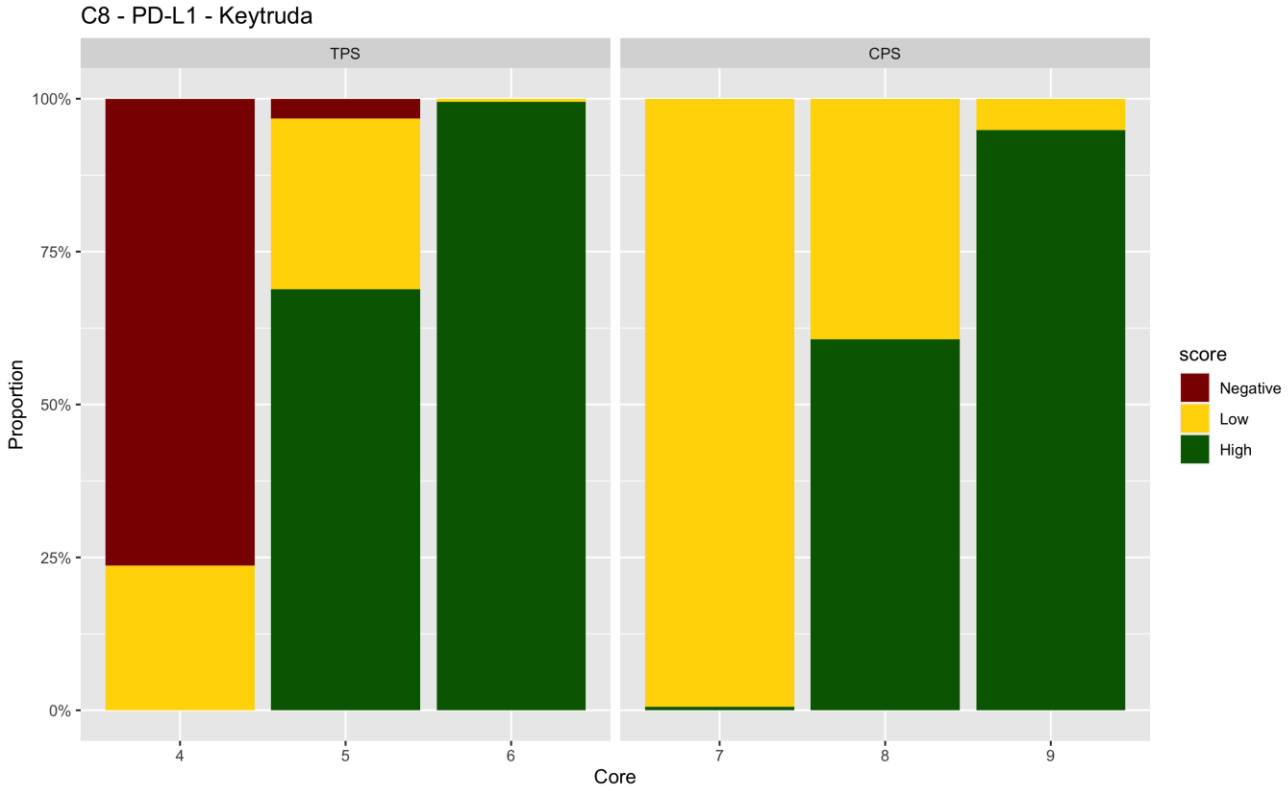
The mAb clone 22C3 was the most widely used concentrated Ab within a LD assay (n=40) providing a pass rate of 73% (29 of 40), 22% optimal.
25 laboratories used the mAb clone 22C3 on the Ventana BenchMark stainer platform and obtained a pass rate of 84% (21 of 25). The performance on BOND III / BOND MAX was shown to be inferior, as a pass rate of 20% was seen (1 of 5), no optimal. The Ab was applied by similar central protocols settings on BOND compared to both BenchMark and Omnis, but with limited success.

Other concentrated Abs as e.g. clones CAL10 and ZR3 could also be used to generate sufficient and optimal results within LD assays.

RTU Abs without predictive claims are now categorized separately in Table 1 and includes the Ventana rmAb clone SP263, 790-4905 previously being listed as CE IVD /FDA approved PD-L1 assays. The RTU format is similar to the two other Ventana SP263 products but positioned as an open "analytical product" for PD-L1 and option to validate the protocol by the end-user. Laboratories using the RTU format as recommended by Ventana obtained a pass rate and proportion of optimal results comparable to the "locked PD-L1 assays" 741-4905 and 740-4907 (see Table 1).

PD-L1 interpretation and scoring consensus:

Participants were asked to score each of the cores using either tumour proportion score (TPS) for the NSCLCs or combined positive score (CPS) for the urothelial carcinomas.



Graph 1. **NordiQC PD-L1 run C8: Tumour Proportion scores (TPS) in NSCLCs (core no. 4-6) and Combined Positive Score (CPS) in urothelial carcinomas (core no. 7-9).**

As seen in Graph 1, a relatively high consensus rates were observed for the tissue core 4,6,7 and 9, whereas the consensus rates were significantly lower in tissue core 5 and 8.

For the tissue core no. 5, tissue heterogeneity was observed throughout the blocks and the TPS category changed between TPS low and TPS high. This also seems to be reflected in the scoring results from the participants. Concerning tissue core no 8, which only had positive immune cells (and not any positive tumour cells), a relative high number of laboratories scored this as CPS low. One possible explanation

could be that some laboratories scored the core using the TPS scoring system and not the CPS scoring system, which also includes positive immune cells in the CPS score.

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

Controls

Tonsil and placenta were used as positive and negative tissue controls. In this and previous assessments, tonsil was found to be superior to placenta, as tonsil displayed a dynamic and clinical relevant range of PD-L1 expression levels, whereas placenta virtually only contained cells (throphoblasts) with high level PD-L1 expression.

In tonsil, protocols with optimal results typically provided the following reaction pattern:

A moderate to strong predominantly membranous staining reaction in dispersed crypt epithelial cells, a weak to moderate, typically punctuated membranous staining reaction of the majority of germinal centre macrophages and scattered interfollicular lymphocytes and macrophages. No staining reaction in the vast majority of lymphocytes and normal stratified squamous epithelial cells.

It was observed that rmAb SP263 (741-4905, 790-4905/4907, Ventana) typically provided a higher proportion of positive inter- and intra-follicular immune cells compared to the Dako/Agilent 22C3 PD-L1 assays (SK006 and GE006). For other clones, e.g. mAb clone CAL10 typically a stronger staining reaction in more germinal centre macrophages were seen compared to mAb clone 22C3, when the clones provided otherwise optimal and accurate results in the carcinomas. This emphasizes that the expected test performance characteristics in tonsil must be correlated to the PD-L1 IHC test/clone used both for inter- and intra PD-L1 IHC reproducibility evaluation.

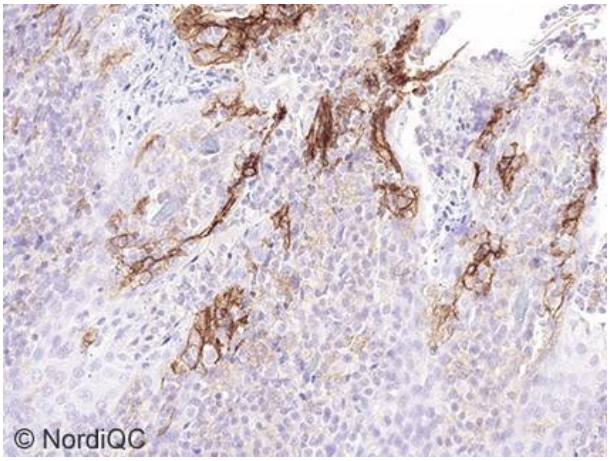


Fig. 1a
Optimal staining result of tonsil using the PD-L1 IHC 22C3 pharmDx kit, GE006, Dako/Agilent on Omnis following the recommended protocol settings.
Dispersed epithelial crypt cells show a moderate to strong predominantly membranous staining reaction, while normal stratified squamous epyhelial cells are are negative.
Also compare with Figs. 2a – 4a, same protocol.

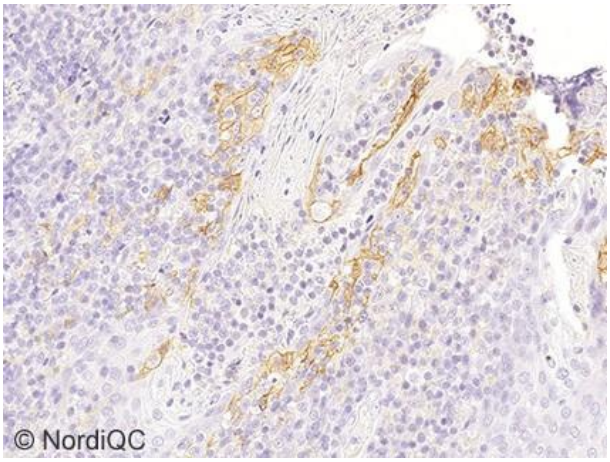


Fig. 1b
Staining result of tonsil, using the rmAb clone CAL10 within a laboratory developed test for PD-L1.
The epithelial crypt cells are clearly demonstrated, but the overall result using this protocol provided an insufficient result, as shown in Figs. 2b and 4b, same protocol.
The epithelial crypt cells express high levels of PD-L1 and cannot reliably be used to evaluate the accuracy and level of analytical sensitivity for PD-L1 IHC.

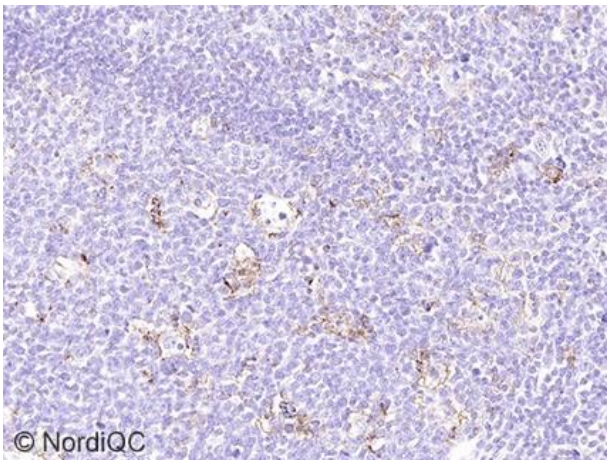


Fig. 2a
Optimal staining result of tonsil using the same protocol as in Fig. 1a
A weak to moderate, but distinct punctuated membranous staining reaction of germinal centre macrophages and dispersed lymphocytes is seen.
No staining reaction in the vast majority of lymphocytes and a high signal-to-noise ratio is observed.

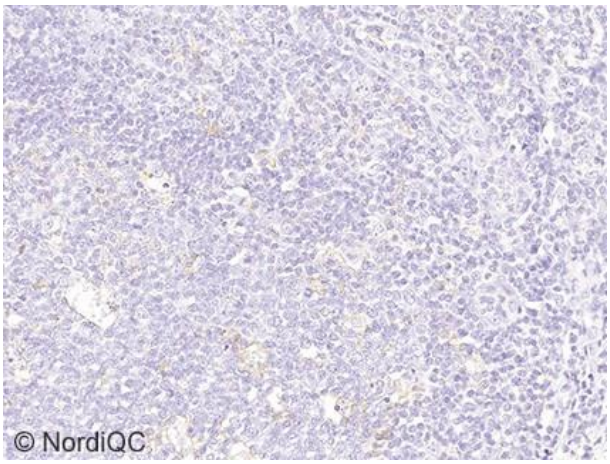


Fig. 2b
Insufficient staining result of tonsil using same protocol as in Fig. 1b.
Germinal centre macrophages and lymphocytes are virtually negative indicating a reduced analytical sensitivity of the IHC protocol applied.
Also compare with Fig. 4b and the diagnostic impact of the low level analytical sensitivity in a NSCLC.

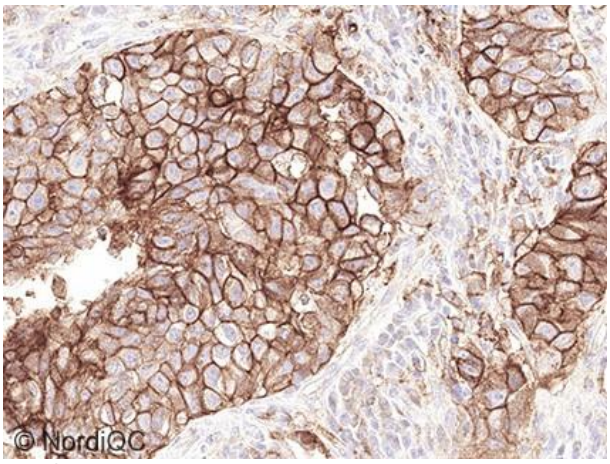


Fig. 3a
Optimal staining result of the NSCLC, tissue core no. 6, using the same protocol as in Figs. 1a and 2a.
Virtually all tumour cells show a moderate to strong membranous staining reaction.
The tumour was categorized as TPS high (≥50%) and thus eligible for first line immune therapy with KEYTRUDA® (different regional cut-offs occur).

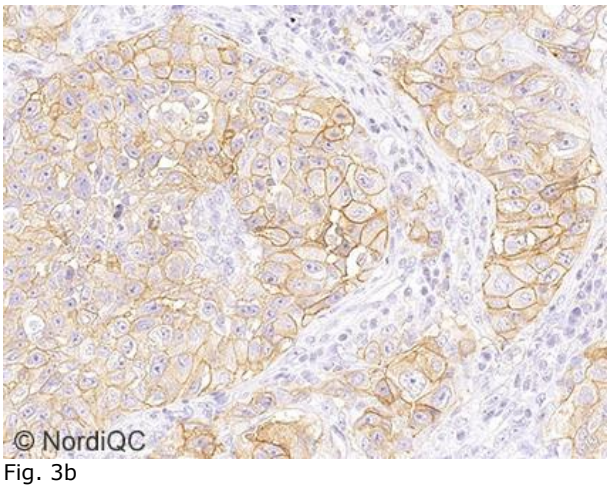


Fig. 3b
Staining result of the NSCLC, tissue core no. 6, using the same protocols as in Figs. 1b and 2b.
The vast majority of tumour cells show a weak to moderate membranous staining reaction.
The tumour was categorized as TPS high (≥50%) and thus eligible for first line immune therapy with KEYTRUDA® (different regional cut-offs occur).
However also compare with Fig. 4b, same protocol.

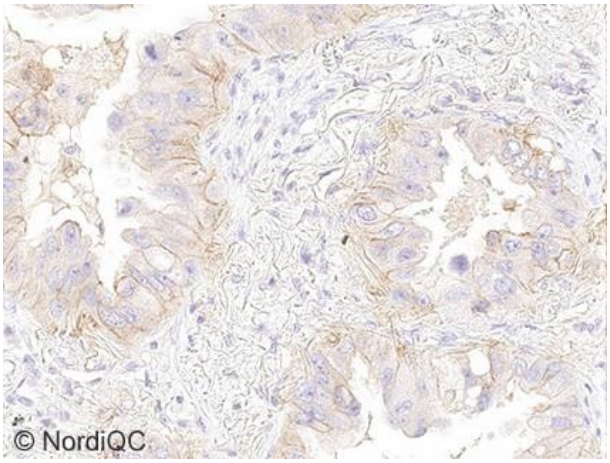


Fig. 4a
Optimal staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1a - 3a.
Approximately 20-30% of tumour cells show a weak to moderate partial membranous staining reaction.
The tumour was categorized as TPS low (≥1-49%) and thus eligible for second line immune therapy with KEYTRUDA® (different regional cut-offs occur).

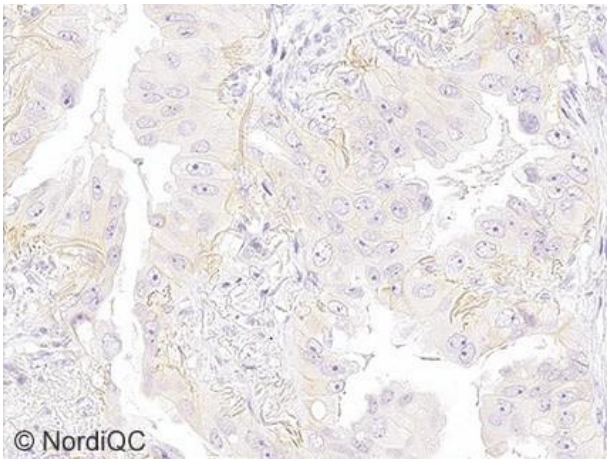


Fig. 4b
Insufficient staining result of the NSCLC, tissue core no. 5, using the same protocol as in Figs. 1b – 3b.
No or <1% of tumour cells show a membranous staining reaction changing the TPS category from the expected low to neg.

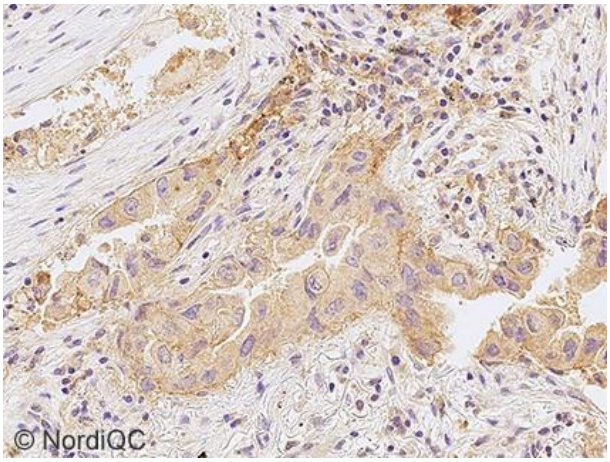


Fig. 5a
Insufficient staining result of the NSCLC, tissue core no. 5, using the rmAb CAL10 by inappropriate protocol settings providing a poor-signal-to-noise ratio compromising the read-out.
The PD-L1 status cannot reliable be determined due to an excessive cytoplasmic staining reaction.
Also see Fig. 5b, same protocol.

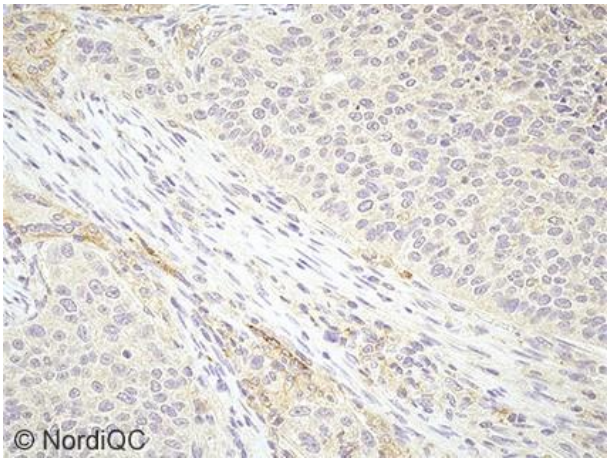


Fig. 5b
Insufficient staining result of the NSCLC, tissue core no. 4, using the rmAb CAL10 by inappropriate protocol settings providing a poor-signal-to-noise ratio compromising the read-out.
This tumour was TPS neg as defined by the NordiQC reference standard methods and virtually all participants.
The excessive cytoplasmic staining reaction obscures a reliable evaluation of PD-L1 reaction in the membranes.

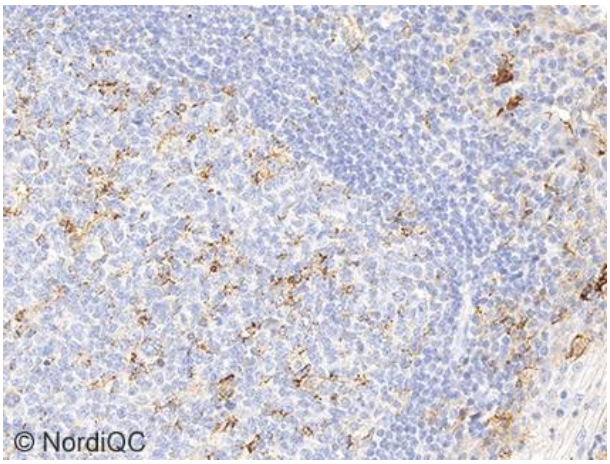


Fig. 6a
Optimal staining result in tonsil using the CDx assay SP263, 741-4905 Ventana/Roche on BenchMark Ultra following the recommended protocol settings.
An increased number of immune cells both within the germinal centres and in the interfollicular areas is seen as compared to the level typically seen for the Dako/Agilent PD-L1 IHC 22C3, pharmDx assays. Compare with Fig. 1a.
Despite the increased number of normal immune cells demonstrated, the SP263 assay was successful and accurate in the 4 urothelial carcinomas included in this run (however SP263 is not approved for PD-L1 status in urothelial carcinoma using CPS for stratification with KEYTRUDA®).

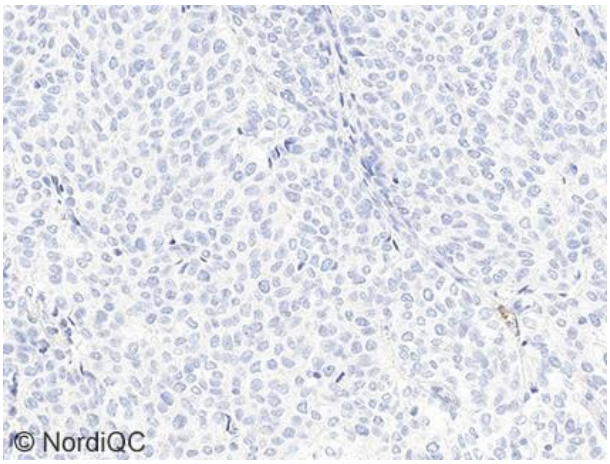


Fig. 6b
Optimal staining result in the urothelial carcinoma, tissue core no. 7, using the CDx assay SP263, 741-4905 Ventana/Roche on BenchMark Ultra following the recommended protocol settings.
The PD-L1 status concordant to level expected and as defined by the NordiQC reference method PD-L1 IHC 22C3 pharmDx kit, SK006, Dako/Agilent on AS48, as CPS<10.
It has to be mentioned that the Ventana SP263 is not approved for PD-L1 status in urothelial carcinoma using CPS for stratification with KEYTRUDA®.

SN/LE/RR 14.12.2020