

Lessons Learned, Challenges Taken, and Actions Made for “Precision” Immunohistochemistry. Analysis and Perspectives From the NordiQC Proficiency Testing Program

Søren Nielsen, BMS,* Michael Bzorek, BMS,† Mogens Vyberg, MD,‡
and Rasmus Røge, MD, PhD*

Abstract: Immunohistochemistry (IHC) has for decades been an integrated method within pathology applied to gain diagnostic, prognostic, and predictive information. However, the multimodality of the analytical phase of IHC is a challenge to ensure the reproducibility of IHC, which has been documented by external quality assessment (EQA) programs for many biomarkers. More than 600 laboratories participate in the Nordic immunohistochemical Quality Control EQA program for IHC. In the period, 2017-2021, 65 different biomarkers were assessed and a total of 31,967 results were evaluated. An overall pass rate of 79% was obtained being an improvement compared with 71% for the period, 2003-2015. The pass rates for established predictive biomarkers (estrogen receptor, progesterone receptor, and HER2) for breast carcinoma were most successful showing mean pass rates of 89% to 92%. Diagnostic IHC biomarkers as PAX8, SOX10, and different cytokeratins showed a wide spectrum of pass rates ranging from 37% to 95%, mean level of 75%, and attributed to central parameters as access to sensitive and specific antibodies but also related to purpose of the IHC test and validation performed accordingly to this. Seven new diagnostic biomarkers were introduced, and all showed inferior pass rates compared with the average level for diagnostic biomarkers emphasizing the challenge to optimize, validate, and implement new IHC biomarkers. Nordic immunohistochemical Quality Control operates by “Fit-For-Purpose” EQA principles and for programmed death-ligand 1, 2 segments are offered aligned to the “3-dimensional” approach—bridging diagnostic tests, drugs to be

offered, and diseases addressed. Mean pass rates of 65% and 79% was obtained in the 2 segments for programmed death-ligand 1.

Key Words: immunohistochemistry, external quality control, precision medicine

(*Appl Immunohistochem Mol Morphol* 2023;31:452–458)

Immunohistochemistry (IHC) is an indispensable assay being consistently performed in anatomic pathology primarily for the subclassification of neoplasms.^{1,2} IHC is basically a descriptive assay aimed to determine whether a target analyte is present or absent, and accordingly the result classified as either positive or negative.^{3,4} For a minor number of analytes, such as human epidermal growth factor receptor 2 (HER2) and programmed death-ligand 1 (PD-L1), a “semiquantitative” result is obtained. IHC serves as a diagnostic, prognostic, and predictive assay, and is, in the era of precision medicine, increasingly applied as an established biomarker method for decision-making of personalized targeted therapies.⁵ Because of the essential value, standardization of IHC to ensure high-level test reproducibility has been in focus in the last decades. The journey from a “special stain” to an “in-situ proteomic” biomarker method is in particular based on the perspectives by Clive Taylor in the publication “*The total test approach to standardization of immunohistochemistry*” in 2000.⁶ IHC is a complex method, and the end result is influenced by multiple parameters in the preanalytic, analytic, and postanalytic phases.^{7,8} Millions of different protocols can be generated for each biomarker with great risk of poor reproducibility compromising patient safety.⁹ Because of the complexity of IHC methods, it is impossible to standardize these to 1 universal methodology. The focus should rather be directed on reproducibility of IHC results and the 2 most central tools for laboratories to assure high-quality and reproducible IHC testing are based on internal quality management adhering to national and international standards and participation in external quality assessment (EQA) programs.^{10–12}

This publication will describe the working principles of the NordiQC (Nordic immunohistochemical quality

Received for publication March 7, 2022; accepted September 9, 2022.

From the *NordiQC, Department of Pathology, Aalborg University Hospital, Aalborg; †Department of Surgical Pathology, Zealand University Hospital, Roskilde; and ‡Center for RNA Medicine, Aalborg University, Copenhagen, Denmark.

The authors declare no conflict of interest.

Reprints: Søren Nielsen, BMS, NordiQC, Institute of Pathology, Aalborg University Hospital, Ladegårdsgade 3, Aalborg 9000, Denmark (e-mail: sn@rn.dk).

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

control) program, focusing on central observations generated since the latest publications of NordiQC data^{13,14} and integrate the lessons learned from the 4-paper series for IHC quality assurance published by Torlakovic and colleagues and Cheung and colleagues with particular emphasis on the need for “Fit-For-Purpose IHC” relevant for all stakeholders within IHC.^{4,15–17} At present, about 600 laboratories from >60 countries participate in the NordiQC program, which allow a detailed analysis of the lessons learned, challenges taken, and actions made to offer EQA for IHC in the era of precision diagnostics.

MATERIALS AND METHODS

NordiQC offers 3 IHC modules: (1) a general module for type 1 diagnostic IHC tests^{4,18} relevant to identify and subclassify neoplasms; (2) a breast cancer module focusing on type 2 IHC tests^{4,18} for HER2, estrogen receptor (ER), and progesterone receptor (PR); and (3) a companion diagnostic module focusing on PD-L1.

In brief, the working principles for NordiQC are based on the distribution of unstained slides from tissue-microarrays to the participants, performing the analysis with their standard IHC method for the respective biomarkers. The stained slides are returned to NordiQC for central evaluation and compared with a “designated true value” enabling an objective assessment of the performance conducted by an “expert panel” composed of pathologists and biomedical scientists. Subsequently, feedback to the participants is given through publicly available assessment-reports and specific assessment marks to the individual participants with tailored recommendation for method improvement when needed. The concept is mainly focused on an evaluation of the basic level of analytical sensitivity and specificity of the methods used. The 5 core elements of the working procedures by NordiQC are shown in Figure 1.

The NordiQC “reference standard IHC method” is used as comparator for the evaluation of the participants results. As this method will designate and provide input on the “true” level of expected results, this or these methods must be carefully selected and validated appropriately. NordiQC is located at the Department of Pathology, Aalborg University Hospital, Denmark, and has access to

IHC methods developed, validated accordingly to ISO 15189, and implemented for diagnostics with historic data on results. For type 2 IHC biomarkers as PD-L1 and HER2, commercially available and vendor-validated companion diagnostic (CDx) IHC assays are applied as reference standard methods.

In diagnostic IHC, there is at present no access to reference standards, calibrators, or traceable units to verify the performance of IHC methods.¹⁹ In contrast, these components are applicable in clinical biochemistry to verify the accuracy and precision of methods and serve as core element to assure test reproducibility both for laboratories, EQA programs, and industry-producing reagents and instruments for the analysis. In IHC, immunohistochemical critical assay performance controls (iCAPCs) serve as “pseudo reference materials” used to evaluate the test performance characteristics concerning analytical sensitivity with emphasis on low limit of detection (LLOD), basic analytical specificity and test reproducibility.^{17,20,21} For virtually, all biomarkers evaluated by NordiQC an attempt to identify, characterize, and use relevant iCAPCs have been established. At <http://www.nordiqc.org>, an online library for iCAPCs is available for numerous biomarkers with recommendations on tissue types, purposes, and descriptions of expected results. In this context, it must be mentioned that for certain biomarkers, no tissue with reliable iCAPC capabilities can be identified to evaluate LLOD and IHC test reproducibility.

In addition to the use of iCAPCs, selected patient samples are included in the NordiQC EQA material. The samples are identified accordingly to the purpose of the IHC test and contain the clinical-relevant and diagnostic-relevant expression levels of the biomarker.

Both patient samples and “reference standard materials”/iCAPCs are formalin fixed and paraffin embedded (FFPE) material being processed accordingly to international standards and recommendations of e.g. American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP),²² clinical and laboratory standard institute (CLSI).²³

The central element defining the selection of reference standard IHC method, “reference standard materials”/iCAPCs, patient samples, and read-out criteria for calling a result positive or negative is based on the purpose of the IHC test requiring that the life cycle of IHC from validation process to usage of quality controls must always follow the Fit-For-Purpose concept.^{4,24,25} This is now addressed for all biomarkers assessed by NordiQC and especially central for type 2 biomarkers as PD-L1, where no universal purpose, reference standard method, or read-out criteria exist. Only by offering a “3D” EQA design integrating validated Diagnostic tests as reference methods, focusing on clinical relevant Diseases, and using read-out criteria validated for Drug efficiency, the analytical quality of the participants IHC results can reliably be evaluated.²⁶ As a consequence, 2 segments for PD-L1 are available: 1 segment—PD-L1 TPS/CPS (tumor proportion score/combined positive score) focusing on PD-L1 in, for example, non-small cell lung carcinoma (NSCLC)

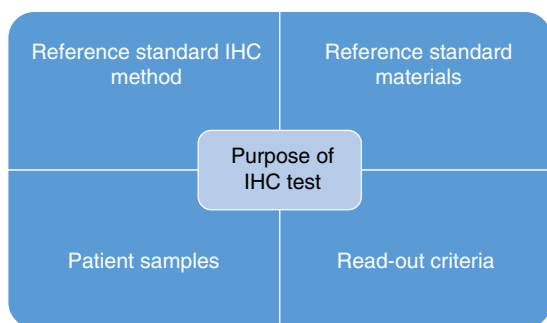


FIGURE 1. The 5 core elements of Nordic immunohistochemical Quality Control external quality assessment analysis. IHC indicates immunohistochemistry.

for treatment with pembrolizumab, and 1 segment—PD-L1 IC (tumor-infiltrating immune cell area) score focusing on PD-L1 for in, for example, triple-negative breast carcinoma for treatment with atezolizumab. As mentioned, the Fit-For-Purpose objective is addressed for both type 2 and type 1 biomarkers, and for the latter typically the central purpose being identification and subclassification of cancer of unknown primary origin and consequently relevant differential diagnostic samples are included in the EQA material.

Each participant's results are assessed by consensus of an expert panel and scored as optimal, good, borderline, or poor. The scores are primarily based on the level of basic analytical sensitivity and specificity of the result, but also related to the technical quality as morphology, signal-to-noise ratio, counterstaining, etc. See details in Table 1.

RESULTS

This section will include selected representative NordiQC results primarily generated in the period from 2017 to 2021. Supplemental data of all NordiQC results is publicly available on <http://www.nordiqc.org>.

Overall, 65 different biomarkers were assessed in the period. In the general module focusing on type 1 IHC tests, the biomarkers were typically only tested 1 time. Selected biomarkers, for example, PAX8, were repeated up to 4 times. The decision to repeat type 1 IHC tests was primarily related to low pass rates but also new purpose(s) of the IHC tests. For the type 2 IHC assays PD-L1, ER, and HER2, these were tested biannually. See Table 2 for all biomarkers evaluated.

A total of 31,967 slides were evaluated by NordiQC in the period 2017-2021. The overall assessment scores are summarized in Figure 2. Seventy-nine percentage of the results were evaluated as sufficient (optimal and good) and 21% as insufficient (borderline and poor). The overall pass rate was improved compared with the data observed in the period from 2003 to 2015, in which an overall pass rate of 71% was reported.^{13,14}

The breast cancer module was the largest module with 324 to 394 participants enrolled for the biomarkers included: ER, PR, and HER2. As shown in Figure 2, the proportion of sufficient and optimal results was higher in this module compared with the levels seen in the general and CDx modules. From the initial assessments within the breast cancer

module, a consistent harmonization of the protocols used by the participants has been observed. This is illustrated in Table 3 focusing on the protocols used for ER. A clear trend toward the usage of full automation, Ready-To-Use (RTU) primary antibodies (Abs) in combination with heat-induced epitope retrieval in an alkaline buffer and multimer-based/polymer-based detection systems.

Similar to the observations for ER, harmonization of the protocols applied for HER2 IHC has emerged with the vast majority of participants using Food and Drug Administration/Conformité Européenne In Vitro Diagnostic (FDA/CE-IVD)-approved assays on the expense of laboratory developed tests (LDTs). In the latest run B32 2021, 78% of the participants used FDA/CE-IVD-approved HER2 IHC CDx assays, whereas 22% used LDTs based on a concentrated primary Ab or a RTU Ab without predictive claim. The CDx HER2 assays have been more successful providing superior pass rates compared with LDTs in all assessments. Among the different commercially available CDx assays for HER2 IHC, PATHWAY (Roche) was found to be most accurate proving a mean pass rate of 96% (range: 86 to 100%) compared with a mean pass rate of 76% (range: 56% to 94%) for LDTs.

A mean pass rate of 89%, 92%, and 90% was obtained for ER, PR, and HER2 IHC, respectively. For all the 3 biomarkers, false-negative results have been the main feature of insufficient results but also false-positive results and inadequate technical quality of the IHC result compromising the read-out, were seen (Table 4).

In relation to the improved performance of the results within the breast cancer module, a change in the application and use of tissue controls has been effectuated by the participants. In run B1 2006, 12% of the participants applied on-slide controls for the ER-slides submitted to NordiQC compared with 48% in run B32, 2021. This change is concordant to the recommendations from ASCO/CAP for ER and PR testing²¹ and recommended by the international ad hoc expert committee for IHC.²⁰

For the results generated within the general module for 60 different biomarkers, a mean pass rate of 75% was seen, being the lowest for the 3 modules. The pass rate should be seen in the light of the complexity and diversity of biomarkers included in this module, as many new biomarkers are introduced and for well-established targets,

TABLE 1. NordiQC Scoring Criteria

Score	Scoring Criteria	Scoring Status
Optimal	The staining is considered perfect or close to perfect in all of the included tissues.	Sufficient: passed
Good	The staining is considered fully acceptable in all of the included tissues. However, the protocol may be optimized to improve the proportion and/or staining intensity of cells demonstrated and/or optimize the technical quality as signal-to-noise ratio, morphology, etc.	Sufficient: passed
Borderline	The staining is considered insufficient, because of a generally too weak staining reaction, a false-negative/false-positive result in one of the included tissues and/or inferior technical quality. The protocol should be optimized.	Insufficient: failed
Poor	The staining is considered insufficient, because of a false-negative/false-positive result in 2* or more of the included tissues and/or inferior technical quality. The protocol should be optimized.	Insufficient: failed

*For type 2 IHC assays, 1 false-negative/false-positive result can be scored as poor.

TABLE 2. Biomarkers Assessed in NordiQC 2017-2021

General Module			Breast Module	Companion Module
ALK-LUNG (2)	CEA	OCT3/4	ER (10)	PD-L1 TPS/CPS (10)
AMACR	CGA	p16	HER2 (10)	PD-L1 IC score (5)
ASMA (2)	CK19	p40	PR (5)	—
Bcl2	CK20	p53	—	—
Bcl6	CK5 (2)	p63	—	—
BRAF	CK7	PAX8 (4)	—	—
BSAP	CK-LMW (2)	PMS2 (2)	—	—
CD10	CK-PAN (2)	Podop	—	—
CD117 (2)	CMYC	S100 (2)	—	—
CD15	CR	S100	—	—
CD23	ECAD	SATB2	—	—
CD30	EPCAM	SMAD4	—	—
CD31 (2)	ERG	SMH	—	—
CD45	GATA3 (2)	SOX10 (2)	—	—
CD5	MLA (3)	SYP	—	—
CD56	MLH1 (2)	TdT	—	—
CD68	MSH2 (2)	TTF1	—	—
CD79a	MSH6 (2)	UP	—	—
CD8	MUM1	VIM	—	—
CDX2	NKX3.1 (2)	WT1	—	—

Number in brackets indicate number of repeats.
TPS/CPS indicates tumor proportion score/combines positive score.

new purposes and associated new demands for the IHC protocols required. In the period, 7 new biomarkers were introduced and all gave an inferior pass rate compared with the average level in the module as shown in Table 5.

The top 3 most successful biomarkers were CD3, cytokeratin 7 (CK7), and CK20 with pass rates of 95%, 94%, and 94%, respectively. SOX10 showed the most substantial improvement as the pass rate changed from 45% in the first assessment to 92% in the latest run. PAX8 showed a consistent low pass rate in the 4 runs conducted, mean pass rate 45%, range 37% to 56%. The assessment for p53 in run 63, 2021 gave a significant inferior pass rate of 46% compared with 79% in the previous run 38, 2013.

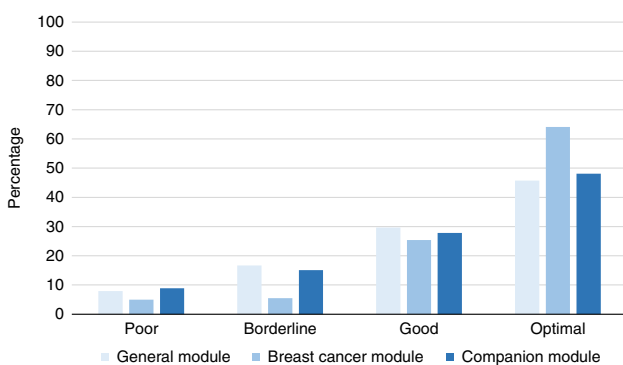


FIGURE 2. Assessment scores in the 3 Nordic immunohistochemical quality control immunohistochemistry modules 2017-2021.

TABLE 3. Harmonization of Protocols and Pass Rates for ER Among NordiQC Participants

Run No.	8	B1	B15	B32
Year	2003	2006	2013	2021
Ready-To-Use antibody (%)	17	20	66	88
Alkaline buffer for HIER (%)	75	85	94	96
Multimer/polymer detection system (%)	61	71	93	99
Fully automated IHC platform (%)	4	24	59	89
Pass rate (%)	50	75	77	89

HIER indicates heat-induced epitope retrieval.

The overall results obtained in the CDx module for PD-L1 TPS/CPS initiated in 2017 were very similar to the results for similar predictive and semiquantitative type 2 IHC assays as ER and HER2, when they were introduced in the NordiQC program and in detail described in previous NordiQC publications^{13,14}. As shown in Figure 3, a pass rate of 50% was seen in the first C1-run for PD-L1 TPS/CPS and after consecutive repeats with possibility for the participants to adjust the protocols, the pass rate gradually improved in the 9 successive assessments.

Similar to the data for HER2, FDA/CE-IVD-approved CDx assays for PD-L1 TPS/CPS were more successful than LDTs. For LDTs, the mean pass rate in the 10 runs was 71% (range 20% to 91%), compared with 91% for CDx assays (range: 80% to 95%). The CDx assays comprised 22C3—SK006/GE006 (Agilent), 28-8—SK005 (Agilent), and SP263—741-4905 (Roche) while LDTs based on concentrated or RTU Abs without predictive claim.

In the CDx module, the segment for PD-L1 IC score comprised 5 assessments from 2019 to 2021, with a relatively consistent pass rate; a mean value of 65% (range: 55% to 76%) was achieved. In contrast to the TPS/CPS segment, indicating an interchangeability of different PD-L1 CDx assays and identification of “best practice” and successful LDTs, the CDx assay SP142, 741-4860 (Roche) outperformed both other PD-L1 assays as 22C3 SK006/GE006 (Agilent) and SP263 790-4905 (Roche), as well as LDTs. Cumulated data for the 5 assessments for PD-L1 IC score revealed a mean pass rate of 88% (range: 78% to 93%) for the CDx assay SP142, 741-4860 (Roche), whereas other PD-L1 CDx

TABLE 4. Pass Rates and Features of Insufficient Results for ER, PR, and HER2 IHC. NordiQC, 2017-2021

	Insufficient Results			Mean Pass Rate, n (%)
	FN (%)	FP (%)	TQ (%)	
ER	86	5	9	89 (70-94)
PR	50	40	10	92 (85-99)
HER2	63	12	25	90 (76-97)

Parentheses indicate range in the assessments performed.

*FN: false negative, FP: false positive, TQ: technical quality.

CDx indicates companion diagnostic assays; LDTs, laboratory developed tests.

TABLE 5. New Biomarkers Introduced in NordiQC, 2017-2021

Target	Pass Rate (%)	Participants
BRAF	72	135
C-MYC	57	172
ERG	67	130
NKX3.1	65	49
SATB2	58	105
SMAD4	42	52
URO II/III	45	66

assays and LDTs provided a mean pass rate of 20% (range 0% to 74%).

The insufficient results in the segment for PD-L1 TPS/CPS have mainly been characterized by false-negative test results, whereas insufficient results in the PD-L1 IC score segment most frequently caused by read-out challenges related to excessive staining reaction of tumor cells compromising the scoring of PD-L1 in immune cells (Table 6).

DISCUSSION

Concordant with previously published NordiQC data,^{13,14} the current data for 2017-2021 emphasizes that laboratories still face challenges to implement and validate/verify IHC methods. The overall pass rate of 79% was improved compared with the level of 71% obtained in the period 2003-2015, which is encouraging, but in the time of precision medicine the need to further ensure IHC test accuracy and precision is warranted. The improvement may be attributed to several parameters as enhanced internal quality management, access to refined products/instrumentation for IHC, and participation in EQA programs and publications providing all stakeholders within IHC guidelines how to optimize and validate/verify IHC methods.^{4,15-17}

A certain harmonization and consolidation of IHC “best practice” methodology has emerged as shown for ER (Table 3). In all the three NordiQC modules, a clear trend toward the use of full automation, application of RTU Abs, CDx assays and general use of an IHC

protocol backbone based on efficient heat-induced epitope retrieval, and sensitive/specific detection systems. Especially, the application of CDx assays used in compliance with instructions for use has been imperative for the high and consistent pass rates for both HER2 and PD-L1 and shown to be a clear asset for precision testing for precision medicine for these 2 biomarkers. Per se, the analytical part of IHC within pathology is transitioning from LDT based to a “black box kit testing” approach similar to other laboratory disciplines as clinical chemistry with all the pros and cons associated. The central element for both LDTs and IHC kits is anchored on the Fit-For-Purpose concept and if an IHC biomarker test is developed accurately with a clear intended purpose, this facilitates the process. For well-established biomarkers with a relatively narrow purpose as CD3, CD79a, CK7, CK20, and SOX10, pass rates of 92% to 95% were obtained. In contrast, a significant inferior performance and pass rates were achieved for biomarkers with multiple diagnostic purposes as Pan-CK, CK5, and Melan A. As example, the assessment run 49, 2017 for Melan A with focus on demonstration of this target analyte in both melanoma and sex cord tumors revealed a pass rate of 60%, whereas a pass rate of 88% was obtained in run 60, 2020 focussing only on melanomas. The inferior pass rate in the assessment including sex cord tumors was related to LLOD being much lower in this entity compared with the level needed for melanomas and the methods applied by the participants were mainly calibrated for lesions with high expression levels and not adequately optimized for the LLOD in sex cord tumors.

In this context, refined needs and read-out criteria of existing IHC biomarkers might influence the performance of IHC methods. In this context, IHC for p53 was historically a biomarker for TP53 mutations demonstrating a nuclear overexpression of the p53 protein caused by missense mutation and, for example, accordingly used in Barret esophagus.²⁷ Recently, it was discovered that for gynecological pathology, other TP53 mutations occurred causing absence/null expression of nuclear p53 protein or p53 protein mainly being localized in the cytoplasmic compartment.^{28,29} In run 63, 2021, the NordiQC assessment for p53 addressed the use of IHC for p53 in tubo-ovarian carcinomas and included samples with both missense mutations causing p53 overexpression and mutations causing p53 absence. The pass rate declined significantly to 46% in

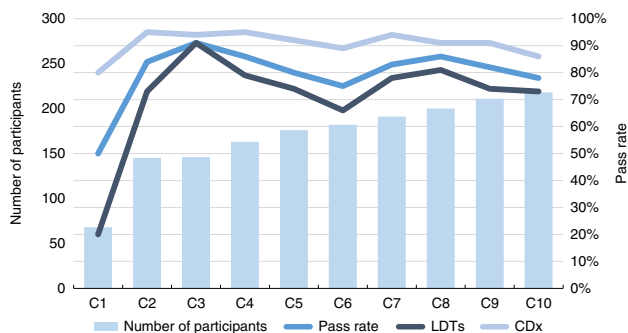


FIGURE 3. Pass rates for programmed death-ligand 1 tumor proportion score/combined positive score in Nordic immunohistochemical quality control, 2017-2021. CDx indicates companion diagnostic assays; LDTs, laboratory developed tests.

TABLE 6. Pass Rates and Features of Insufficient Results for PD-L1. NordiQC 2017-2021

	Insufficient Results			Mean Pass Rate, n (%)
	FN (%)	FP (%)	TQ (%)	
PD-L1 TPS/CPS	74	10	16	79 (50-91)
PD-L1 IC score	36	6	58	65 (60-76)

Parentheses indicate range in the assessments performed.

FN indicates false negative; FP, false positive; IC score, tumor-infiltrating immune cell area score; TPS/CPS, tumor proportion score/combined positive score; TQ, technical quality.

contrast to 79% in the previous run focusing only on p53 nuclear overexpression. The results clearly indicated that the IHC methods both LDTs and RTU kits were mainly developed for p53 overexpression, but not calibrated to identify mutations causing absence/null expression and hereby not applicable as accurate predictor of TP53 mutations and surrogate for molecular testing and revalidation warranted.

The access to and application of sensitive and specific primary Abs is essential for the quality of IHC for any biomarker, and EQA data can reveal which Abs/clones for a given biomarker to be used to give reliable results and also indicate which to be avoided. This was in particular relevant for SOX10, in which a pass rate of 45% was obtained in the initial run 45, where polyclonal Abs (pAbs) gave a pass rate of 14% in contrast to monoclonal Abs (mAbs) giving a pass rate of 73%. In the recent run 60, 2020 for SOX10 a pass rate of 92% was seen and a contributing factor being the extended use of mAbs by 98% of the participants on the expense of pAbs being used by 42% in the first run. The lack of access to reliable Abs have shown to be an obstacle to improve the performance of certain biomarkers assessed. For PAX8, and despite maintaining the same purpose of the IHC test and being repeated 4 times in the period 2017-2021, virtually no improvement of the results for PAX8 have been made. At present, only few specific PAX8 Abs are available, whereas most vendors offer PAX8 Abs cross-reacting with PAX2, PAX5, and PAX6 causing mischaracterization and lack of reproducibility in the testing and reporting of PAX8 in tumors.^{30,31} Both the NordiQC data and several studies have revealed that the apparent PAX8 expression in many tumors as thymomas, lymphomas, breast carcinoma, and medullary thyroid carcinomas is caused by cross-reaction of certain Abs to especially PAX5 in B cells and PAX6 in neuroendocrine cells.^{32,33}

A central factor impacting the overall pass rate of an IHC biomarker is also related to the robustness of the commercial available antibodies and performance on the different IHC instruments or platforms. For the successful biomarkers CD3, CD79a, CK7 and CK20 several antibodies are available providing the expected performance on both semi- and fully-automated IHC platforms, whereas for e.g. Melan-A and PAX8 the most widely used antibodies have an inferior performance on most commonly used fully-automated platforms.³¹

For virtually all biomarkers evaluated in NordiQC, it was found possible to develop IHC methods with accurate analytical capabilities. However, it is not only the goal to develop and validate the methods, but there is also a huge need to identify tools to monitor the reproducibility of the results. In this aspect, the use of iCAPCs is valuable and applicable for many biomarkers, and a great number of normal tissues can be used as descriptive control for LLOD, for example, hepatocytes for Pan-CK,²⁰ follicular dendritic cells for ER,^{21,34} etc. In order to aid laboratories to monitor the reproducibility and consistency of their IHC methods, NordiQC has designed an online “map of tissue controls” being available at www.nordiqc.org, with scanned slides and descriptions of expected results in the tissues to verify LLOD

and basic specificity of almost 100 IHC biomarkers. However, for many important “semiquantitative” biomarkers as HER2 and PD-L1, no reliable iCAPCs or alternative documented controls have been identified. One of the emerging tools with potential to be “the missing brick” might be the calibrators developed by Steve Bogen¹⁹ and associates. In short, the concept is based on several microbead pellets coated with a range of different concentrations of the target analyte enabling the option to measure the LLOD of a given IHC test and then correlate this to the relevant LLOD as determined by a validated IHC test. This will in theory induce a tool or calibrators to objectively evaluate IHC test reproducibility and can be implemented in the entire IHC life cycle from development, validation, and final method transfer for routine diagnostics. Analogue to microbeads serving as reference standard materials or calibrators for IHC, studies to use cell lines in combination with image analysis show promising potential as standardization tools for IHC.^{35,36} The identification of suitable cell lines, characterization of critical expected levels of a certain biomarker by validated IHC assays in the cell lines, and an objective image analysis algorithm will without question be huge asset to monitor IHC test reproducibility both by laboratories and EQA programs.

The working principles for NordiQC and IHC quality assessment have till now been based on expert evaluation conducted by a group of assessors evaluating the performance of the the participants results. This process is and has been effective for many diagnostic and predictive IHC biomarkers, but also shown to be laborious and in addition a more objective and granular evaluation for especially semi-quantitative biomarkers as ER, HER2 and PD-L1 is warranted. The use of DIA has been documented to be a promising tool for reproducible and objective biomarker scoring as alternative to the manual biomarker scoring methods and might be of high value also within EQA programs to document and measure the accuracy of submitted IHC biomarker results comparing these objectively to the expected and critical reference levels separating positive versus negative results by applying the different relevant thresholds.³⁷ However, the use of DIA both for diagnostic use and for EQA must be validated and followed by meticulous quality control when replacing existing scoring methods.³⁸

The quotation by Clive Taylor back in 2000 “Immunohistochemistry is technically complex, and no aspect of this complexity can be ignored, from the moment of collecting the specimen to issuance of the final report” is still after 22 years highly valid and been confirmed by the lessons learned, challenges taken, and actions made in the NordiQC EQA program for IHC. We still need further progress in the standardization of IHC, but especially for present predictive biomarkers based on IHC, the analytical aspect is now on a high qualitative level. With this in place and potential of the rapidly growing and evolving area of digital analysis and artificial intelligence to aid pathologist in the read-out, IHC will also in the next decades be a central element in precision diagnostics.

REFERENCES

- Bellizzi AM. An algorithmic immunohistochemical approach to define tumor type and assign site of origin. *Adv Anat Pathol*. 2020; 27:114–163.
- Stelow EB, Yaziji H. Immunohistochemistry, carcinomas of unknown primary, and incidence rates. *Semin Diagn Pathol*. 2018; 35:143–152.
- Taylor CR. Predictive biomarkers and companion diagnostics. The future of immunohistochemistry: “in situ proteomics,” or just a “stain”? *Appl Immunohistochem Mol Morphol*. 2014;22:555–561.
- Cheung CC, D’Arrigo C, Dietel M, et al. Evolution of quality assurance for clinical immunohistochemistry in the era of precision medicine: part 1: fit-for-purpose approach to classification of clinical immunohistochemistry biomarkers. *Appl Immunohistochem Mol Morphol*. 2017;25:4–11.
- Gatalica Z, Feldman R, Vranić S, et al. Immunohistochemistry-enabled precision medicine. *Cancer Treat Res*. 2019;178:111–135.
- Taylor CR. The total test approach to standardization of immunohistochemistry. *Arch Pathol Lab Med*. 2000;124:945–951.
- Kim SW, Roh J, Park CS. Immunohistochemistry for pathologists: protocols, pitfalls, and tips. *J Pathol Transl Med*. 2016;50:411–418.
- Hecke DV. Routine immunohistochemical staining today: choices to make, challenges to take. *J Histotechnol*. 2002;25:45–54.
- Education Guide. Talor CR, Rudbeck L. *Immunohistochemical Staining Methods*, 6th edChapter 4:46-59. Dako. Available at: https://www.agilent.com/cs/library/technicaloverviews/public/08002_ihc_staining_methods.pdf. Accessed January 26, 2022.
- Schneider F, Maurer C, Friedberg RC. International Organization for Standardization (ISO) 15189. *Ann Lab Med*. 2017;37:365–370.
- Guzel O, Guner EI. ISO 15189 accreditation: requirements for quality and competence of medical laboratories, experience of laboratory. *I Clin Biochem*. 2009;42:274–278.
- Dufraing K, Fenizia F, Torlakovic E, et al. Biomarker testing in oncology—requirements for organizing external quality assessment programs to improve the performance of laboratory testing: revision of an expert opinion paper on behalf of IQNPath ABSL. *Virchows Arch*. 2021;478:553–565.
- Nielsen S. External quality assessment for immunohistochemistry: experiences from NordiQC. *Biotech Histochem*. 2015;90:331–340.
- Vyberg M, Nielsen S. Proficiency testing in immunohistochemistry—experiences from Nordic Immunohistochemical Quality Control (NordiQC). *Virchows Arch*. 2016;468:19–29.
- Torlakovic EE, Cheung CC, D’Arrigo C, et al. Evolution of quality assurance for clinical immunohistochemistry in the era of precision medicine—part 2: immunohistochemistry test performance characteristics. *Appl Immunohistochem Mol Morphol*. 2017;25:79–85.
- Torlakovic EE, Cheung CC, D’Arrigo C, et al. Evolution of quality assurance for clinical immunohistochemistry in the era of precision medicine. Part 3: technical validation of immunohistochemistry (IHC) assays in clinical IHC laboratories. *Appl Immunohistochem Mol Morphol*. 2017;25:151–159.
- Cheung CC, D’Arrigo C, Dietel M, et al. Evolution of quality assurance for clinical immunohistochemistry in the era of precision medicine: part 4: tissue tools for quality assurance in immunohistochemistry. *Appl Immunohistochem Mol Morphol*. 2017;25:227–230.
- Canadian Association of Pathologists-Association canadienne des pathologistes National Standards Committee, Torlakovic EE, Riddell R, et al. Canadian Association of Pathologists-Association canadienne des pathologistes National Standards Committee/Immunohistochemistry: best practice recommendations for standardization of immunohistochemistry tests. *Am J Clin Pathol*. 2010;133:354–365.
- Bogen SA. A root cause analysis into the high error rate in clinical immunohistochemistry. *Appl Immunohistochem Mol Morphol*. 2019; 27:329–338.
- Torlakovic EE, Nielsen S, Francis G, et al. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. *Appl Immunohistochem Mol Morphol*. 2015;23:1–18.
- Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J Clin Oncol*. 2020;38:1346–1366.
- Wolff AC, Hammond MEH, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *Arch Pathol Lab Med*. 2018;142: 1364–1382.
- Clinical and Laboratory Standard Institute. I/LA28 Quality Assurance for Design Control and Implementation of Immunohistochemistry Assays, 2nd Edition. Available at: <http://www.clsi.org>.
- Torlakovic EE. Fit-for-purpose immunohistochemical biomarkers. *Endocr Pathol*. 2018;29:199–205.
- Swanson PE. Validation and proficiency testing of biomarker immunohistochemistry: diagnostic accuracy “Fit for Purpose” in a 3D world. *Appl Immunohistochem Mol Morphol*. 2019;27:247–250.
- Cheung CC, Barnes P, Bigras G, et al. Fit-For-Purpose PD-L1 biomarker testing for patient selection in immuno-oncology: guidelines for clinical laboratories from the Canadian Association of Pathologists-Association Canadienne Des Pathologistes (CAP-ACP). *Appl Immunohistochem Mol Morphol*. 2019;27:699–714.
- Keswani RN, Noffsinger A, Waxman I, et al. Clinical use of p53 in Barrett’s esophagus. *Cancer Epidemiol Biomarkers Prev*. 2006;15: 1243–1249.
- Köbel M, Ronnett BM, Singh N, et al. Interpretation of P53 immunohistochemistry in endometrial carcinomas: toward increased reproducibility. *Int J Gynecol Pathol*. 2019;38Suppl 1(Iss 1 suppl 1): S123–S131.
- Köbel M, Piskorz AM, Lee S, et al. “Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma.” *J Pathol Clin Res*. 2016;2:247–258.
- Røge R, Nielsen O, Bzorek M, et al. NordiQC assessments of PAX8 immunoassays. *Appl Immunohistochem Mol Morphol*. 2018; 26:221–224.
- Gucer H, Caliskan S, Kefeli M, et al. Do you know the details of your PAX8 antibody? Monoclonal PAX8 (MRQ-50) is not expressed in a series of 45 medullary thyroid carcinomas. *Endocr Pathol*. 2020;31:33–38.
- Kilgore MR, Bosch DE, Adamson KH, et al. Unexpected PAX8 immunoreactivity in metastatic high-grade breast cancer. *Appl Immunohistochem Mol Morphol*. 2019;27:637–643.
- Gown AM. Diagnostic immunohistochemistry: what can go wrong and how to prevent it. *Arch Pathol Lab Med*. 2016;140:893–898.
- Sapino A, Cassoni P, Ferrero E, et al. Estrogen receptor alpha is a novel marker expressed by follicular dendritic cells in lymph nodes and tumor-associated lymphoid infiltrates. *Am J Pathol*. 2003;163: 1313–1320.
- Aung TN, Acs B, Warrell J, et al. A new tool for technical standardization of the Ki67 immunohistochemical assay. *Mod Pathol*. 2021;34:1261–1270.
- Lanng MB, Møller CB, Andersen AH, et al. Quality assessment of Ki67 staining using cell line proliferation index and stain intensity features. *Cytometry A*. 2019;95:381–388.
- Stålhammar G, Fuentes Martinez N, Lippert M, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. *Mod Pathol*. 2016;29:318–329.
- Bui MM, Riben MW, Allison KH, et al. Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer: Guideline From the College of American Pathologists. *Arch Pathol Lab Med*. 2019;143:1180–1195.