Review paper

Review of a decade of IHC quality – expectations, disappointments, and challenges in commonly used assays based on NordiQC experience

Kristi Bøgh Anderson, Søren Nielsen

NordiQC, Department of Pathology, Aalborg University Hospital, Aalborg, Denmark

Immunohistochemistry (IHC) has undergone substantial expansion during the past decades, with an increasing number of assays becoming essential to routine diagnostic, prognostic, and predictive pathology. Parallel to this development, external quality assessment (EQA) programs have accumulated extensive experience regarding the real-world performance of commonly used IHC assays across laboratories. This review summarises observations derived from the Nordic Immunohistochemical Quality Control Program and its EQA activities over the last 10 years, focusing on expectations, recurring disappointments, and emerging challenges in IHC testing. Particular attention is given to the analytical and diagnostic accuracy and quality of IHC assays being impacted by protocol variability and the need to align IHC methods to evolving clinical purposes. The accumulated experience demonstrates that while overall pass rates are acceptable for many markers, significant and recurrent limitations persist, particularly when assays are repurposed or when analytical sensitivity requirements change. These findings underline the continued need for critical assay validation, identification and application of appropriate controls, and evidence-based standardisation of IHC in the era of precision diagnostics.

Key words: immunohistochemistry, external quality assessment, NordiQC, assay performance, precision diagnostics.

Introduction

Immunohistochemistry (IHC) remains a cornerstone of diagnostic surgical pathology, supporting tumour classification, prognostic assessment, and increasingly therapeutic decision-making. Over the past decade, the applicational range of IHC has expanded considerably, driven by advances in targeted therapy, companion diagnostics, and refined biomarker stratification. This expansion has resulted in a growing number of predictive assays being implemented in routine workflows, often under substantial time and resource constraints.

External quality assessment (EQA) programs provide a unique perspective on the real-world performance of IHC assays. External quality assessment testing is an important tool for providing an objective assessment of the technical and analytical competence of participating laboratories. It evaluates the comparability and reproducibility of tests among laboratories and measures performance against established reference values. For IHC assays, real-world data on performance under routine testing conditions, including the selection of antibodies, detection systems, and IHC platforms, can be extracted. This enables the identification of best practice methodologies and supports the harmonisation of testing results across laboratories.

The Nordic Immunohistochemical Quality Control (NordiQC) Program, established in 2003, is

a non-profit EQA organisation that has conducted systematic proficiency testing in IHC for more than two decades [1]. This activity is guided by the overarching mission of improving the quality of IHC and, through this, supporting patient safety as well as accurate diagnostics and treatment selection.

The present review provides a reflective synthesis of NordiQC experiences from approximately the last 10 years. Rather than only aiming to define optimal protocols, the article highlights expectations vs. observed performance, recurrent failure modes, and emerging challenges encountered as IHC adapts to new clinical purposes.

NordiQC external quality assessment: scope and methodology

NordiQC conducts regular proficiency testing cycles covering routine diagnostic markers, breast cancer biomarkers, HER2 *in situ* hybridisation (ISH), and selected companion diagnostic biomarkers. Its non-commercial framework enables independent evaluation of testing performance across IHC assay and product providers, IHC platforms, and laboratories, with publicly available results and structured individual feedback to participating laboratories. Participation in the program is offered on a cost-recovery basis, with fees intended to cover operational expenses without any profit motive. Typically, laboratories receive unstained sections of tissue microarrays with 5–10 different tissue samples covering relevant normal and neoplastic tissue materials, and they perform IHC on these using their routine protocols for the biomarker to be evaluated. Results are assessed by expert evaluators comparing each result submitted against reference slides with predefined evaluation criteria including described expected test performance characteristics. All reference slides and expected results are based on validated reference standard IHC assays. Each result evaluated is graded according to both the technical quality, such as morphology, ease of read-out and counterstaining, and the analytical and diagnostic quality, evaluating whether the assay has been calibrated for the intended purpose. Detailed reports including recommended protocols and recommended tissue controls are made publicly available. In addition, each participating laboratory receives an individual performance assessment expressed as a NordiQC grade (Passed – sufficient; optimal, good. Failed – insufficient; borderline or poor). For results graded other than optimal, the assessment is supplemented with brief expert comment highlighting potential analytical limitations and indicating, based on the applied protocol settings, aspects that could reasonably be adjusted to improve performance [2].

During the last decade, NordiQC has evaluated tens of thousands of IHC stains across different modules. Overall pass rates were around 80%, with failure rates of approximately 20%, showing minor variation between years and modules [2, 3]. Because the data are collected from many laboratories over multiple assessment rounds, they allow identification of recurring patterns rather than isolated mistakes in individual laboratories.

Importantly, NordiQC assessments are designed to evaluate whether IHC assays are calibrated and aligned for their analytical and diagnostic fit-for-purpose performance. Consequently, observed shortcomings often reflect misalignment between the level of sensitivity provided by the assay and the level required for clinical use. In addition to the evaluation of assay calibration, NordiQC also provides assessment of the technical quality of the IHC results. A technically inadequate result, for example caused by severely impaired morphology or a poor signal-to-noise ratio, can have a negative impact on the read-out and thereby compromise the reliability of the testing outcome.

General performance trends and recurrent failure modes

Across markers and modules, the majority of insufficient results were characterised by staining that was too weak or false-negative reactions, accounting for approximately 75–90% of failures. False-positive results constituted a smaller proportion, while severe technical artefacts such as impaired morphology or poor signal-to-noise ratio were relatively uncommon.

These findings underscore that inadequate analytical sensitivity remains the dominant limitation and challenge in routine IHC testing. In many instances, protocols that perform adequately for patient samples with high-expressing levels of the target analyte fail when applied for samples with lower but still clinically relevant expression levels. This is a recurring pattern observed across multiple markers and platforms, highlighting the need to calibrate IHC assays on evidence-based optimisation with attention paid to diagnostic purpose and associated critical range of expression levels rather than having a narrower window typically focusing on high-level expressing samples.

The existence of clone-specific performance issues is also evident. NordiQC data have identified a group of primary antibody clones associated with recurrent poor performance. Such observations emphasise the importance of antibody selection as a critical determinant of assay performance. If a robust, high-affinity, and specific antibody is selected, the IHC method set-up is less critical, and a high-quality result can be obtained using many different settings. In contrast, if a less robust, less sensitive, or less specific antibody is selected, the basis for a reproducible result is compromised and the IHC end result will

typically be of poor quality despite many different and complex protocol settings being tested. In short: garbage in, garbage out [4].

Evolving knowledge and shifting expectations: p53

The p53 IHC assay exemplifies how changing biological understanding directly influences assay performance requirements. Over time, recognition of distinct p53 expression patterns corresponding to different TP53 mutation types has increased expectations regarding both analytical sensitivity and interpretive accuracy.

NordiQC assessments have demonstrated substantial variation in pass rates depending on protocol design and p53 expression patterns to be demonstrated, reflecting the underlying different TP53 mutations. Concentrated antibodies and modified ready-to-use (RTU) protocols have generally outperformed unmodified RTU protocols, particularly in assessments conducted since 2021. Notably, pass rates for RTU systems used strictly according to manufacturer-recommended settings were markedly lower than those achieved with adjusted protocols in recent assessment rounds.

These observations do not imply inherent inferiority of RTU systems but rather reflect the historical context in which currently available RTU p53 assays were developed and validated. At the time of their introduction, p53 IHC was primarily applied to detect overexpression patterns, and assay calibration was therefore optimised for this purpose and for which these assays still perform reliably [3]. Subsequent advances in the understanding of TP53-related biology have, however, identified additional and more subtle yet clinically relevant expression patterns, including complete absence of staining and aberrant cytoplasmic localisation, which correlate with specific molecular alterations [5, 6]. Detection of these patterns places substantially higher demands on analytical sensitivity and dynamic range. As a consequence, present commercially available RTU assays used in a strict plug-and-play manner may fail to capture these expression profiles and require recalibration or protocol adjustment to remain fit for current clinical purposes (Figure 1). The increasing dominance of RTU systems by the NordiQC participating laboratories, combined with declining use of concentrated antibodies, further amplifies the impact of these limitations.

Antibody choice and evidence-based adaptation: SOX10

The SOX10 assessments provide a contrasting example in which evidence-based adaptation led to measurable improvement. NordiQC data demonstrated a clear association between antibody clone and assay performance, with monoclonal antibodies consistently outperforming polyclonal antibodies across multiple runs (Figure 2). In the initial 2 NordiQC assessment runs for SOX10, protocols based on monoclonal antibodies provided an overall pass rate of 74%, compared to only 10% for protocols based on polyclonal antibodies (only focusing on antibody origin and not taking any other parameters into consideration, such as titre, HIER, detection system, IHC platform, *etc*).

Following dissemination of these findings, laboratories increasingly adopted monoclonal SOX10 antibodies, resulting in improved pass rates in subsequent assessments. This experience illustrates that EQA data can support rational decision-making and performance improvement when laboratories are willing to modify established practices [7].

Fit-for-purpose immunohistochemistry: ALK

The concept of fit-for-purpose IHC is central to NordiQC principles and has become increasingly relevant within the IHC community as assays often are repurposed for new clinical contexts. ALK IHC exemplifies this challenge. Originally developed for the detection of high antigen expression in lymphomas, ALK IHC assays are now required to detect substantially lower expression levels associated with eligibility for targeted therapy in non-small cell lung cancer. NordiQC assessments have demonstrated that protocols and antibody clones optimised for one clinical purpose may be inadequate for another, despite acceptable performance in their original context (Figure 3).

Shifting expectations driven by improved performance: Melan A

Melan A illustrates how the availability of assays with improved analytical performance can shift diagnostic expectations. NordiQC internal data and published observations have shown that the widely used A103 clone displays lower diagnostic sensitivity compared with newer Melan A antibody clones in melanoma detection. As higher-performing assays have become available, staining results that were previously considered acceptable are no longer sufficient, reflecting an upward shift in analytical expectations based on what is technically achievable (Figure 4).

Importantly, clone selection for Melan A remains purpose dependent. While newer clones provide superior sensitivity for melanoma detection, A103 retains a specific diagnostic role due to its reactivity in sex cord-stromal tumours, a feature not shared by other clones [8]. This duality highlights that improved performance for one diagnostic area does not necessarily imply universal replacement and reinforces the need to align antibody choice with the intended clinical application.

Together, these examples illustrate that assay validation cannot be static. As clinical applications evolve and assay performance improves, expectations must be reassessed accordingly to ensure that IHC remains fit for its intended purpose.

Therapy-driven redefinition of analytical requirements: HER2

HER2 IHC represents one of the most striking examples of therapy-driven changes in assay expec-

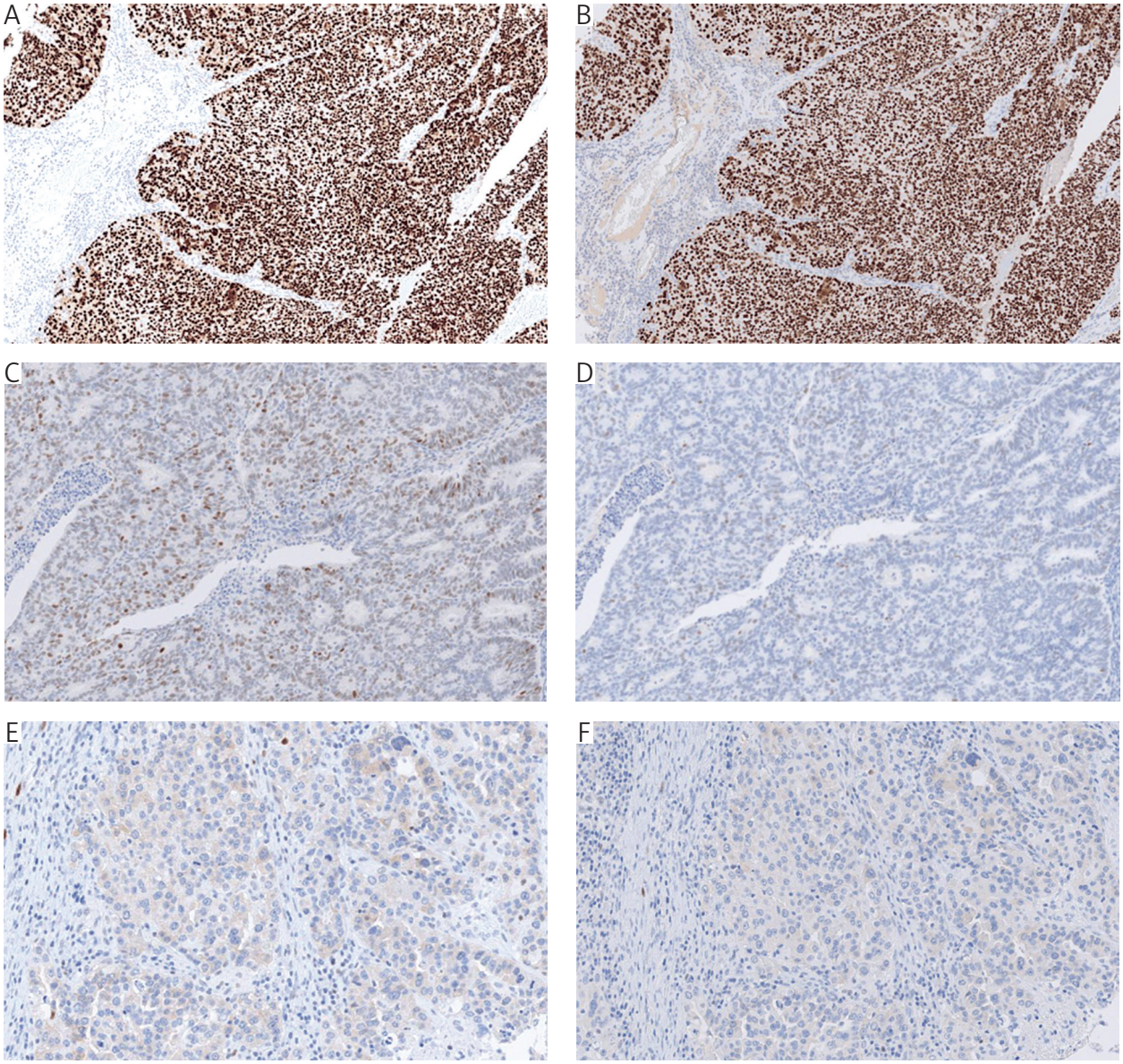

Figure 1. p53 staining reactions in endometrial carcinoma and normal tissue control using two immunohistochemistry assays with different analytical sensitivity Comparison of an optimised, laboratory-modified protocol based on the p53 mouse monoclonal antibody clone DO-7 as ready-to-use (RTU) format using the OptiView detection system (**A, C, E, G, I**) and a vendor-recommended protocol based on the same DO-7 RTU format using the UltraView detection system (**B, D, F, H, J**). Panels **A, C, E, G, I** were stained using the same optimised protocol, while panels **B, D, F, H, J** were stained using the same vendor-recommended protocol. **A, B**) Endometrial carcinoma with NGS-documented TP53 missense mutation. A) Using the optimised protocol, the tumour shows a p53 overexpression pattern with strong, diffuse nuclear staining reaction in the neoplastic cells. **B**) Using the vendor-recommended protocol, a similar p53 overexpression pattern is demonstrated. **C, D**) Endometrial carcinoma with NGS-documented absence of pathogenic TP53 variants (wild-type TP53). **C**) With the optimised protocol, a p53 wild-type staining pattern is observed, characterised by heterogeneous nuclear staining reaction in most tumour cells. **D**) With the vendor-recommended protocol, the staining reaction in tumour cells is markedly reduced, with only a few cells showing very faint nuclear staining reaction. This pattern carries a significant risk of misinterpretation as a p53 absence pattern. **E, F**) Endometrial carcinoma with NGS-documented TP53 mutation affecting the nuclear localisation domain. **E**) Using the optimised protocol, the expected faint cytoplasmic staining reaction for p53 is observed in the neoplastic cells. **F**) Using the vendor-recommended protocol, reduced analytical sensitivity again results in loss of detectable staining reaction, creating a risk that the case would be misclassified as p53 absence type

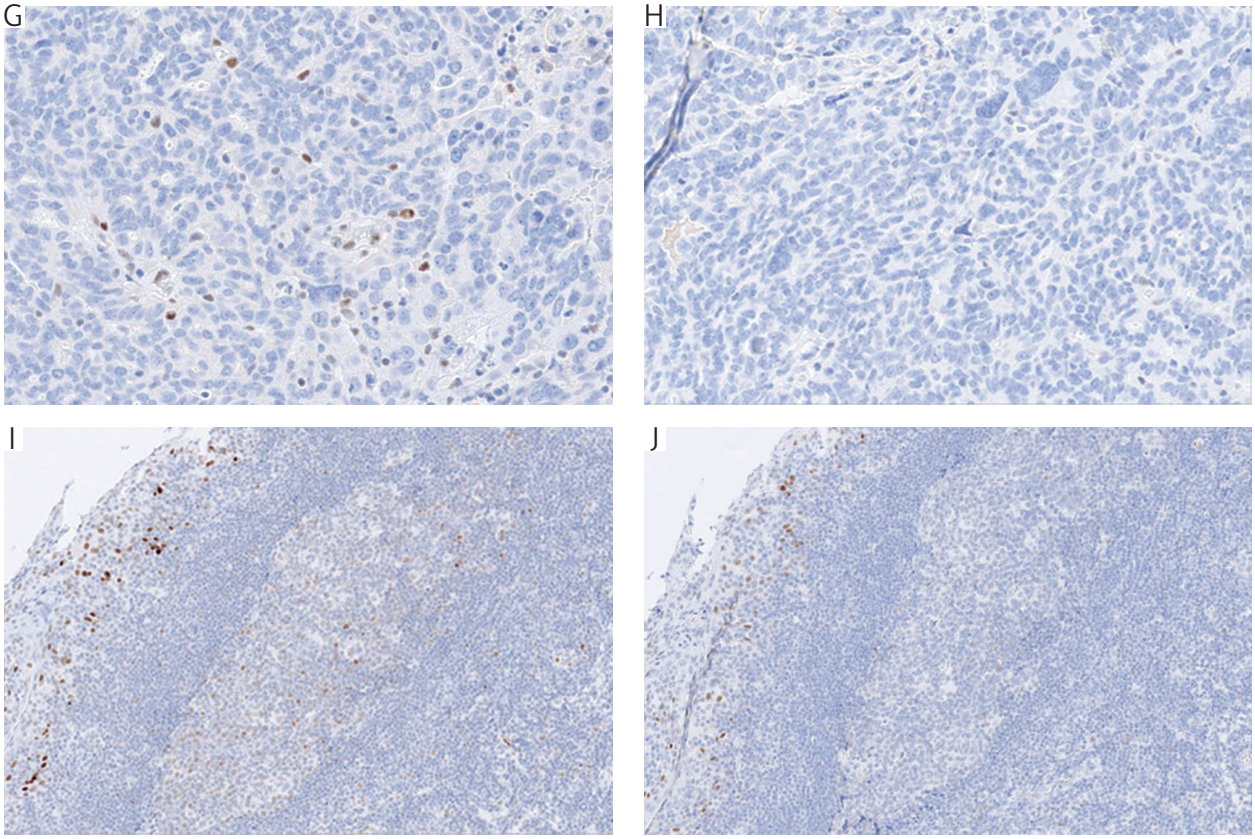

Figure 1. Cont. **G, H**) Endometrial carcinoma with NGS-documented TP53 mutation resulting in a premature stop codon. **G**) With the optimised protocol, complete absence of p53 staining reaction is seen in carcinoma cells, while lymphocytes, stromal cells, and endothelial cells show a wild-type nuclear staining reaction, confirming adequate assay performance through internal positive controls. **H**) With the vendor-recommended protocol, both carcinoma cells and internal control cells are negative. A technical failure of the immunohistochemistry assay or pre-analytical issues cannot be excluded. **I, J**) Tonsil used as normal control tissue. **I**) Using the optimised protocol, a weak to moderate nuclear staining reaction is seen in most germinal centre B-cells, while fewer than 50% of mantle zone B-cells are positive. **J**) Using the vendor-recommended protocol, fewer than 50% of germinal centre B-cells show nuclear p53 staining reaction compared with the optimal result in (**I**), assessed in the same area. Overall, this figure illustrates that reduced analytical sensitivity may still allow recognition of p53 overexpression but substantially compromises correct interpretation of wild-type, cytoplasmic, and absence-type p53 staining patterns

All slides and protocols were performed on the Ventana/Roche BenchMark Ultra platform.

tations. Traditionally, HER2 testing aimed to identify strong overexpression associated with gene amplification. Recent therapeutic developments with antibody-drug conjugates (ADC) have introduced the clinically relevant categories of HER2-low and HER2-ultralow expression [9, 10].

NordiQC assessments have shown that protocols producing acceptable results for classical HER2 overexpression frequently lacked sufficient analytical sensitivity to reliably classify HER2-low tumours. Analyses of assessment data indicate that introduction of HER2-low and HER2-ultralow criteria would markedly reduce overall pass rates, as a substantial proportion of assays that currently pass classical HER2 assessments lack sufficient analytical sensitivity to meet the requirements for HER2-low classification (Figure 5). The inferior pass rate and reduced inter-laboratory reproducibility of HER2 IHC testing for HER2-low and ultralow is without question in part caused by a reduced option to validate or verify the IHC assays for the new purpose. For classical HER2-positive disease, HER2 IHC results can be checked against another validated IHC assay or confirmed by ISH, because HER2 gene amplification is the basis of the classification. This makes IHC assay validation and verification relatively robust.

For HER2-low and HER2-ultralow status, the situation is different. Immunohistochemistry is used to separate HER2-low and HER2-ultralow to be offered ADC treatment from HER2 null not being qualified for the drug. *In situ* hybridisation has limited value because these categories are not defined and separated by amplification status, apart from being without HER2 gene amplification. At present, IHC is therefore the only practical method for identifying these patients. As a result, validation and verification

of IHC assays for HER2-low and HER2-ultralow depend largely on access to reference samples with well-characterised HER2 expression levels.

This limitation likely contributes to the lower pass rates and poorer inter-laboratory reproducibility seen with HER2 IHC testing in the HER2-low and HER2-ultralow setting.

It also highlights an important problem: assays developed for earlier clinical definitions may systematically misclassify cases when applied to these newer categories. Such misclassification may go unnoticed unless EQA materials are specifically designed to represent the lower range of HER2 expressions and are used to confirm that IHC assays are calibrated to detect these clinically relevant levels.

Controls and standardisation challenges

Effective quality assurance depends on the use of appropriate controls that reflect clinically relevant expression levels. NordiQC has emphasised the application of IHC critical assay performance controls (iCAPC) [11] to evaluate the performance of IHC assays concerning analytical sensitivity with empha-

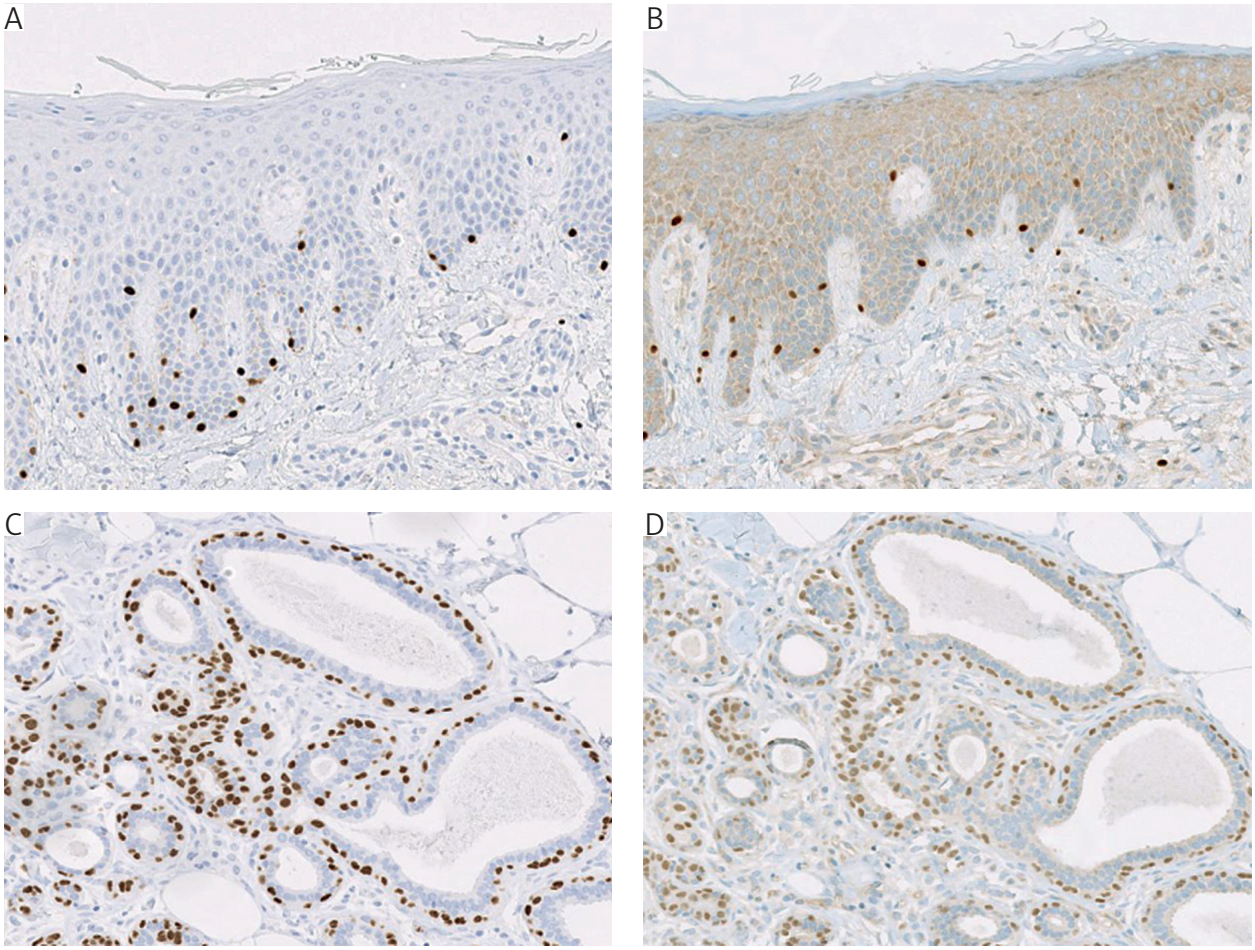

Figure 2. SOX10 staining reactions in normal tissue controls and malignant melanoma using two immunohistochemistry assays with different analytical sensitivity and specificity Comparison of an optimised protocol based on the monoclonal antibody (mAb) clone BC34 (**A, C, E, G, I**) and an insufficient protocol based on a polyclonal antibody (pAb) 383A, Cell Marque (**B, D, F, H, J**). Panels **A, C, E, G, I** were stained using the same optimised protocol, while panels **B, D, F, H, J** were stained using the same insufficient protocol. Optimised protocol: mAb clone BC34 diluted 1:40, heat-induced epitope retrieval (HIER) in CC1 pH 8.5 for 48 min, a 3-step multimer-based detection kit. Insufficient protocol: pAb 383A diluted 1:25, HIER in CC1 pH 8.5 for 32 min, a 3-step multimer-based detection kit. **A, B)** SOX10 staining reaction in skin. **A**) Using the optimised BC34 protocol, virtually all melanocytes show a strong, distinct nuclear staining reaction. No background reaction is observed. **B**) Using the pAb-based protocol in the same field as in (**A**), melanocytes are identified, but aberrant cytoplasmic staining reaction is seen in squamous epithelial cells. **C, D)** SOX10 staining reaction in breast hyperplasia. **C**) With the optimised BC34 protocol, all myoepithelial cells show a moderate to strong nuclear staining reaction, without background staining. **D**) With the insufficient pAb-based protocol in the same field as in (**C**), myoepithelial cells are demonstrated, but the staining intensity is reduced compared with the expected level

The staining pattern observed in **B, D, F, H, J** *was consistently seen with all polyclonal antibodies tested for SOX10. A low affinity for the specific nuclear SOX10 antigen combined with cross-reactivity to a cytoplasmic protein resulted in a low signal-to-noise ratio and insufficient analytical performance.*
All stainings were performed on the Ventana/Roche BenchMark Ultra platform.

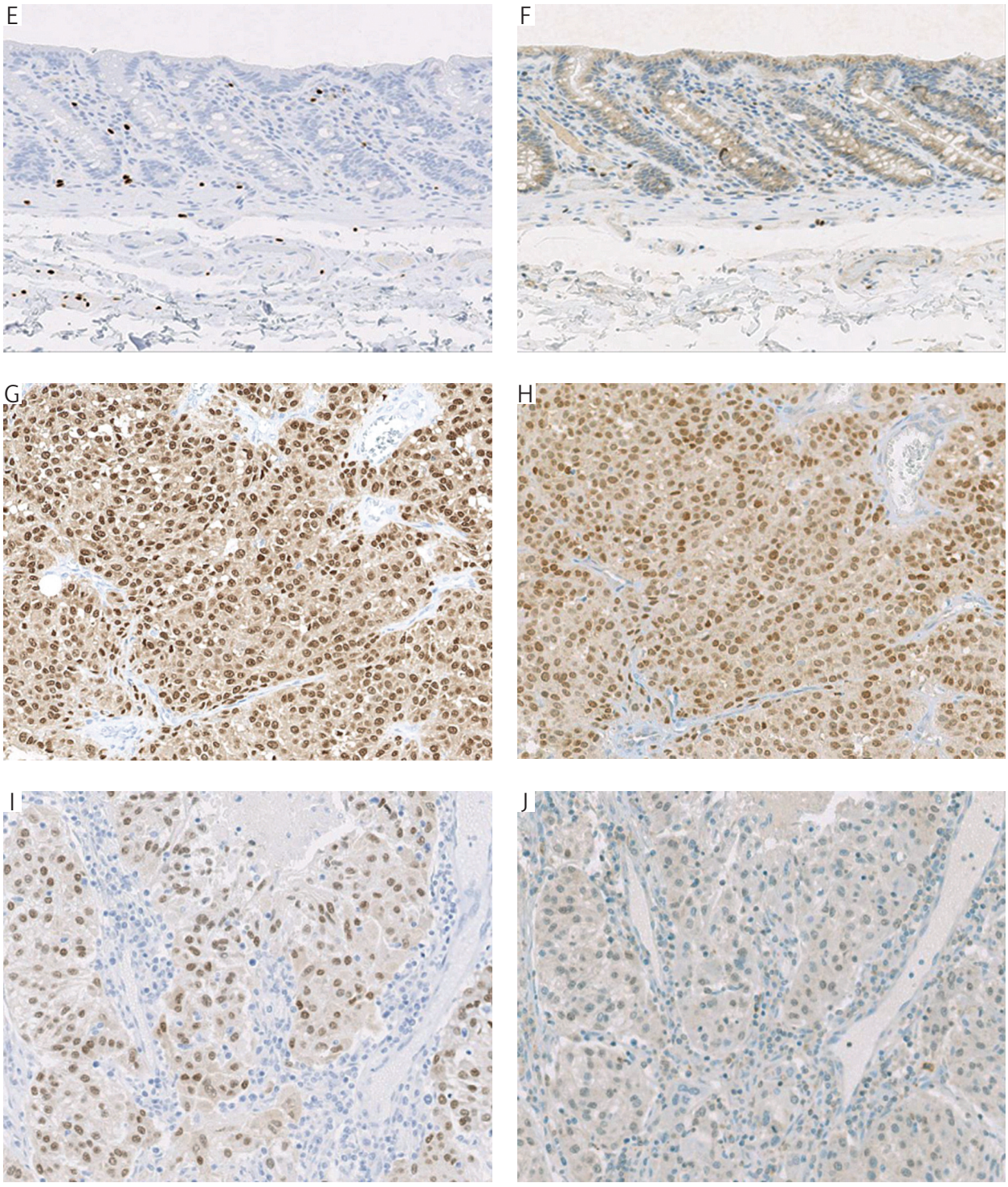

Figure 2. Cont. **E, F)** SOX10 staining reaction in malignant melanoma, high-level expressor. **E**) Using the optimised BC34 protocol, virtually all neoplastic cells show a moderate nuclear staining reaction. A weak cytoplasmic staining reaction is present, but no staining reaction is seen in background or in cells lacking nuclear SOX10 expression. **F**) Using the insufficient pAb-based protocol in the same field as in (**E**), neoplastic cells display only a weak nuclear staining reaction. **G, H)** SOX10 staining reaction in malignant melanoma, low-level expressor. **G**) With the optimised BC34 protocol, the majority of neoplastic cells show a weak to moderate nuclear staining reaction, without background staining. **H**) With the insufficient pAb-based protocol in the same field as in (**G**), only scattered neoplastic cells show a weak and questionable staining reaction. **I, J)** SOX10 staining reaction in colon. **I**) Using the optimised BC34 protocol, Schwann cells show a strong nuclear staining reaction, with no background staining. **J**) Using the insufficient pAb-based protocol, only a weak specific nuclear staining reaction is seen in Schwann cells, accompanied by aberrant cytoplasmic staining reaction in epithelial cells

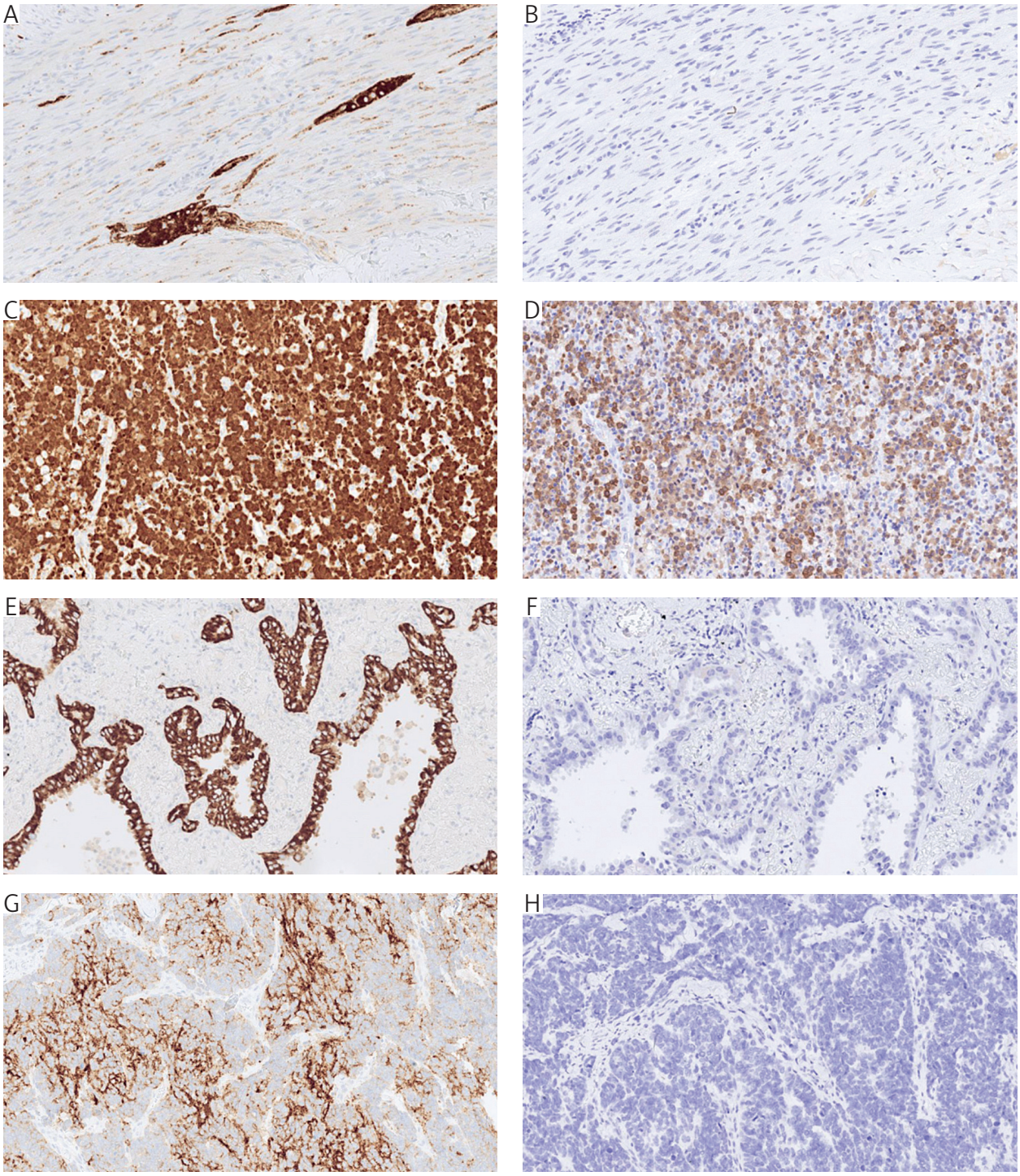

Figure 3. ALK staining reactions in normal tissue control and four different ALK-positive neoplasms using two vendor-recommended ready-to-use immunohistochemistry assays based on different primary antibodies. Comparison of staining performance obtained with the rabbit monoclonal antibody (rmAb) clone D5F3 as ready-to-use (RTU) format (Ventana/Roche, 790-4794) using the Ventana/Roche BenchMark platform (**A, C, E, G**) and the mAb clone ALK1 as RTU format (Dako/Agilent, GA641) using the Dako/Agilent Omnis platform (**B, D, F, H**). All protocols were applied according to vendor-recommended settings. **A, B)** ALK staining reaction in appendix. **A**) Using the D5F3 protocol, optimal staining result is observed in the myenteric plexus, with ganglion cells showing a moderate to strong, distinct cytoplasmic staining reaction and axons displaying a weak to moderate staining reaction. **B**) Using the ALK1 protocol in the same field as in (**A**), both ganglion cells and axons are unstained, demonstrating insufficient analytical sensitivity. **C, D)** ALK staining reaction in anaplastic large cell lymphoma with ALK rearrangement. **C**) With the D5F3 protocol, the neoplastic cells show an intense nuclear and cytoplasmic staining reaction. **D**) With the ALK1 protocol, the neoplastic cells are demonstrated, but both the staining intensity and the proportion of positive cells are markedly reduced compared with (**C**), despite assessment of the same tumour area. **E, F)** ALK staining reaction in lung adenocarcinoma with ALK rearrangement. **E**) Using the D5F3 protocol, most neoplastic cells show a moderate to strong granular cytoplasmic staining reaction, with no background staining. **F**) Using the ALK1 protocol in the same field as in (**E**), the neoplastic cells are false negative. **G, H)** ALK staining reaction in Merkel cell carcinoma. **G**) With the D5F3 protocol, virtually all neoplastic cells show a strong granular cytoplasmic staining reaction, with no background staining. **H**) With the ALK1 protocol in the same field as in (**G**), the neoplastic cells are false negative

sis on low limit of detection (LLOD), basic analytical specificity, and thereby test reproducibility. The tissues and cell populations are particularly selected with special attention to possess descriptive LLOD levels of the specific biomarker to be analysed and consequently sensitive to identify changes in analytical performance of the IHC assay.

NordiQC experience underscores the importance of correct control tissue selection, application, and interpretation, because the use of inadequate controls or controls lacking defined limits of detection may mask the identification of protocols with insufficient analytical sensitivity and thereby lead to unintended clinical consequences.

For many important biomarkers, such as HER2, Claudin 18.2, and PD-L1, no tissue materials with reliable iCAPC characteristics to monitor IHC testing reproducibility have been identified. An emerging tool involves using calibrators consisting of microbead pellets coated with a range of different concentrations of the target analyte [12]. This approach allows for measuring the LLOD of a given IHC assay and correlating it with the reference standard LLOD determined by a validated IHC test. The goal is to create a tool or calibrator set that can objectively

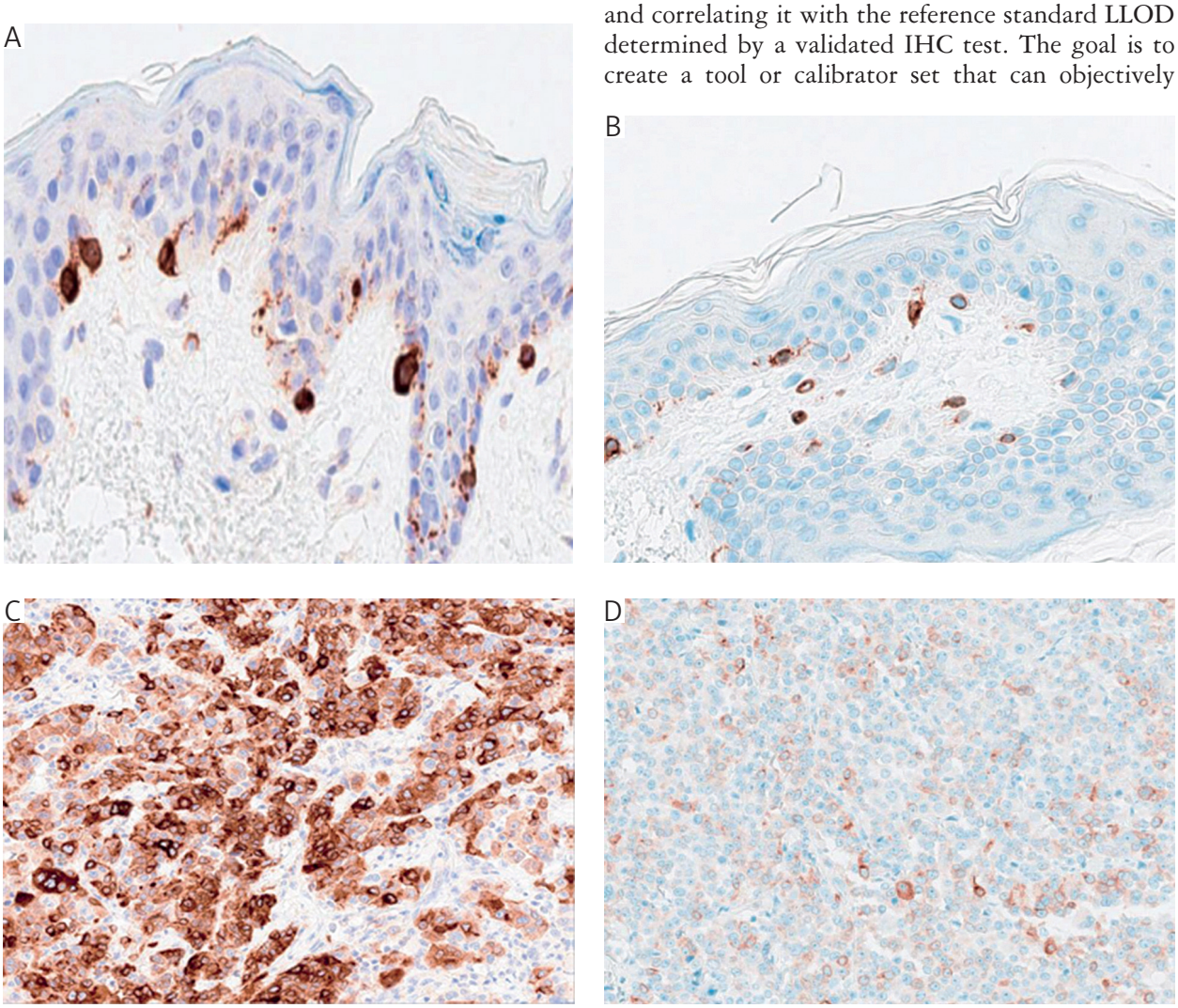

Figure 4. Melan A staining reactions in normal tissue control and malignant melanoma using two immunohistochemistry assays with different analytical sensitivity Comparison of an optimised laboratory protocol based on the monoclonal antibody (mAb) clone BS52 used as a concentrate (**A, C, E, G**) and an insufficient laboratory-developed protocol based on the mAb clone A103 used as a ready-to-use (RTU) format (**B, D, F, H**). Panels **A, C, E, G** were stained using the same optimised protocol, while panels **B, D, F, H** were stained using the same insufficient protocol. **A, B)** Melan A staining reaction in skin. **A**) Using the optimised BS52 protocol, virtually all melanocytes show a strong cytoplasmic staining reaction, with melanocytic dendrites being weak to moderately labelled. **B**) Using the A103-based protocol in the same field as in (A), only dispersed melanocytes are demonstrated, staining intensity is markedly reduced, and melanocytic dendrites are mostly negative. **C, D)** Melan A staining reaction in malignant melanoma demonstrating low expression level. **C**) With the optimised BS52 protocol, virtually all neoplastic cells display a weak to strong cytoplasmic staining reaction. **D**) With the insufficient A103-based protocol, the majority of neoplastic cells are false negative or show only a weak to moderate cytoplasmic staining reaction, compared with (**C**)

Nevertheless, the protocol is considered unreliable due to its low analytical sensitivity, which may risk misdiagnosis of melanomas, as illustrated by comparison of (C, D). Optimised protocol: mAb clone BS52 as concentrate (1:200), efficient HIER) in an alkaline buffer (TRS High, Dako/Agilent) and a 3-step polymer-based detection system
(EnVision Flex+, Dako/Agilent). Insufficient protocol: mAb clone A103 as RTU format (Ventana/Roche, 790-2990) applied as a laboratory-developed assay, HIER in alkaline buffer (CC1, Ventana/Roche) for 48 min, primary antibody incubation for 56 min, and a 3-step multimer-based detection system (OptiView, 760-700, Ventana/Roche).

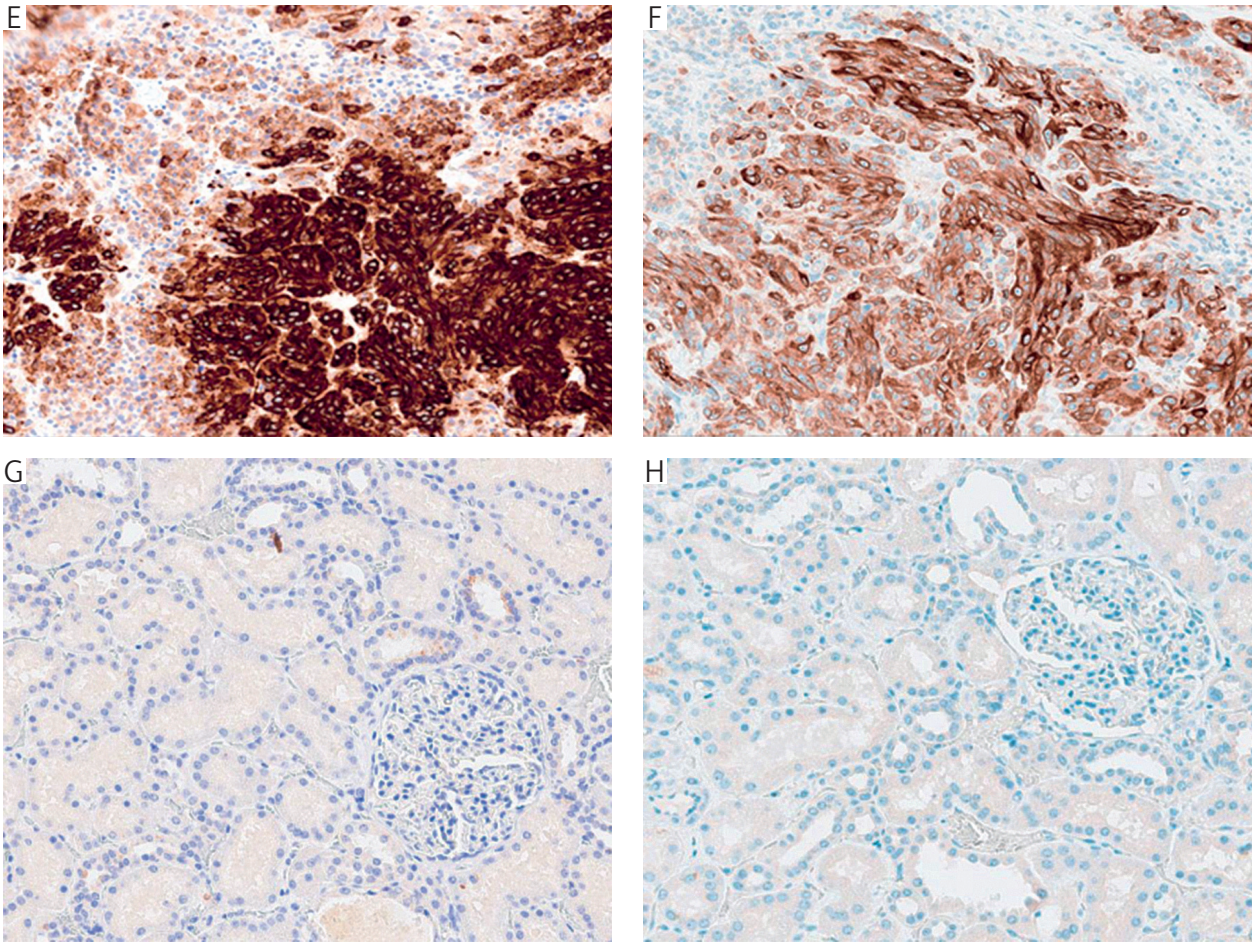

Figure 4. Cont. **E, F)** Melan A staining reaction in malignant melanoma demonstrating high expression level. **E**) Using the optimised BS52 protocol, all neoplastic cells show a strong cytoplasmic staining reaction. **F**) Using the A103-based protocol, most neoplastic cells show weak to strong staining reaction. **G, H)** Melan A staining reaction in kidney. **G**) With the optimised BS52 protocol, tubular epithelial cells show the expected negative staining reaction, demonstrating preserved analytical specificity despite increased sensitivity. **H**) With the A103-based protocol, the tubules show the expected staining pattern

evaluate the reproducibility of IHC assay throughout the entire IHC life cycle, including development, validation, routine diagnostics, and EQA. Like microbeads, cell lines with dynamic and critical clinically relevant expression levels are being increasingly developed to be applied as reliable control material for many IHC assays. Image analysis in combination with suitable cell lines or microbeads have shown promising results to serve as a supplementary tool to evaluate IHC testing reproducibility [13].

Perspective and future directions

Over the past decade, EQA has revealed that IHC remains both robust and vulnerable at the same time. Overall pass rates suggest acceptable general performance, but persistent and recurrent limitations continue to challenge reliable clinical implementation.

Key conceptual advances include recognition of fit-for-purpose assay design, increased awareness of antibody-specific performance, and appreciation of analytical sensitivity. Remaining gaps include insufficient adaptation of protocols to evolving clinical definitions, inconsistent use of controls, and overreliance on predefined RTU settings.

Future priorities should focus on integrating EQA feedback into routine assay management, strengthening evidence-based antibody selection, and ensuring that validation strategies explicitly address emerging clinical needs. Rather than pursuing universal standardisation, realistic progress will require flexible frameworks that accommodate biological complexity. The end goal of IHC standardisation should be to improve and ensure the reproducibility of IHC testing results, while accepting that these may be generated by different IHC assays and protocol settings.

Standardisation efforts are further complicated by the transition from laboratory-developed tests to commercially available RTU systems, driven in part by regulatory changes such as the EU *in vitro* Diagnostic Regulation. While RTU systems offer advantages in terms of reproducibility and workflow integration,

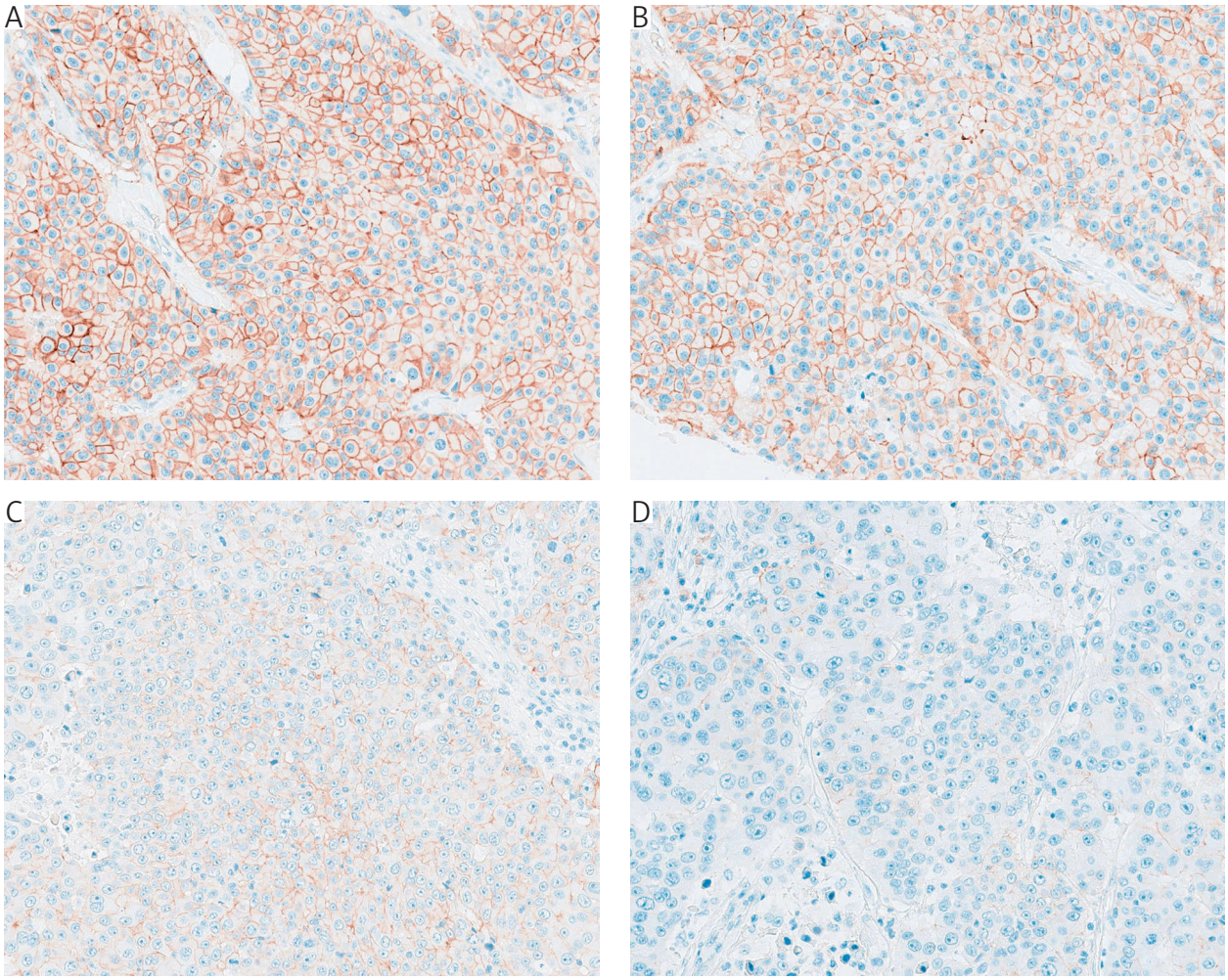

Figure 5. HER2 staining reactions in breast carcinoma illustrating the impact of analytical sensitivity on HER2 classical overexpression and HER2-Low classification Comparison of two immunohistochemistry (IHC) assays for HER2 applied to breast carcinoma tissue cores. **A, C**) Stained using the same validated and approved IHC assay for both HER2 classical and HER2-Low (Ventana/Roche 4B5 RxDx, 790-7167) with adequate analytical sensitivity, while B and D were stained using an IHC with reduced analytical sensitivity. **A, B)** HER2 staining reaction in breast carcinoma with a HER2/chr17 ratio of 3.33. **A**) Expected staining result. More than 10% of the neoplastic cells show a weak to moderate and complete membranous staining reaction, corresponding to a HER2 score of 2+. **B**) Staining result in the same tumour core as in (**A**) using a protocol with reduced analytical sensitivity. More than 10% of the neoplastic cells show a weak but complete membranous staining reaction, still corresponding to a HER2 score of 2+. The proportion of positive cells and staining intensity are reduced compared with (**A**), but the result remains acceptable for classical HER2 overexpression. **C, D)** HER2 staining reaction in breast carcinoma with a HER2/chr17 ratio of 1.35. **C**) Expected staining result as defined by the validated and approved IHC assay for HER2-Low. More than 10% of the neoplastic cells show a weak and partial membranous staining reaction, corresponding to a HER2 score of 1+. **D**) Insufficient staining result. Only very few neoplastic cells show a very faint and partial membranous staining reaction, corresponding to a HER2 score of 0. Both the participant and NordiQC scored the result in (**D**) as 0. Consequently, the tumour was categorised as HER2 0, demonstrating how insufficient analytical sensitivity can lead to misclassification from the expected HER2-Low (1+) status to HER2 0

NordiQC experience indicates that complete protocol closure may not be appropriate for all assay types. In this context, it is also important to emphasise that existing commercially available RTU assays are not consistently adjusted to new diagnostic purposes.

For IHC, we need to start at the end by defining our expected results and then begin the process of transforming an antibody into an IHC assay with a specific intended purpose.

Disclosures

1. Institutional review board statement: Not applicable.
2. Assistance with the article: The authors acknowledge the whole NordiQC team and the external assessors for their continuous contributions to immunohistochemistry quality assessment.
3. Financial support and sponsorship: None.
4. Conflicts of interest: None.

References

1. Vyberg M, Torlakovic E, Seidal T, Risberg B, Helin H, Nielsen S. Nordic immunohistochemical quality control. Croat Med J 2005; 46: 368-371.
2. Nielsen S. External quality assessment for immunohistochemistry: experiences from NordiQC. Biotech Histochem 2015; 90: 331-340.
3. Nielsen S, Bzorek M, Vyberg M, Røge R. Lessons learned, challenges taken, and actions made for "precision" immunohistochemistry. Analysis and perspectives from the NordiQC proficiency testing program. Appl Immunohistochem Mol Morphol AIMM 2023; 31: 452-458.
4. O'Hurley G, Sjöstedt E, Rahman A, Li B, Kampf C, Pontén F, et al. Garbage in, garbage out: a critical evaluation of strategies used for validation of immunohistochemical biomarkers. Mol Oncol 2014; 8: 783-798.
5. Köbel M, Piskorz AM, Lee S, Lui S, LePage C, Marass F, et al. Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. J Pathol Clin Res 2016; 2: 247-258.
6. Bellizzi AM. p53 as exemplar next-generation immunohistochemical marker: a molecularly informed, pattern-based approach, methodological considerations, and pan-cancer diagnostic applications. Appl Immunohistochem Mol Morphol 2023; 31: 507-530.
7. Røge R, Nielsen S, Bzorek M, Vyberg M. NordiQC assessments of SOX10 immunoassays. Appl Immunohistochem Mol Morphol 2017; 25: 377-380.
8. Zuo L, You H, Cai Z, Liao S, Lu X, LiL, et al. Melan-A expression in non-melanocytic carcinoma: a potential diagnostic pitfall. Histol Histopathol 2024; 39: 1037-1041.
9. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med 2022; 387: 9-20.
10. Bardia A, Hu X, Dent R, Yonemori K, Barrios CH, O'Shaughnessy JA, et al. Trastuzumab deruxtecan after endocrine therapy in metastatic breast cancer. N Engl J Med 2024; 391: 2110-2122.
11. Torlakovic EE, Riddell R, Banerjee D, El-Zimaity H, Pilavdzic D, Dawe P, 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.
12. Dabbs DJ, Torlakovic E, Nielsen S, Parry SC, Yu J, Stoos C, et al. New standards in HER2-low testing: the CASI-01 comparative methods study. EBioMedicine 2025; 120: 105919.
13. Lanng MB, Møller CB, Andersen ASH, Pálsdóttir AA, Røge R, Østergaard LR, et al. Quality assessment of Ki67 staining using cell line proliferation index and stain intensity features. Cytom Part A 2019; 95: 381-388.

Address for correspondence

Kristi Bøgh Anderson
NordiQC
Department of Pathology
Aalborg University Hospital
Aalborg, Denmark
e-mail: kristi.anderson@rn.dk