Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Immunohistochemical detection of Ki67 in breast carcinomas

Methology, quality assessment and digital image analysis

Røge, Rasmus

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Røge, R. (2020). *Immunohistochemical detection of Ki67 in breast carcinomas: Methology, quality assessment and digital image analysis.* Aalborg Universitetsforlag. Aalborg Universitet. Det Sundhedsvidenskabelige Fakultet. Ph.D.-Serien

IMMUNOHISTOCHEMICAL DETECTION OF KI67 IN BREAST CARCINOMAS

METHOLOGY, QUALITY ASSESSMENT AND DIGITAL IMAGE ANALYSIS

BY
RASMUS RØGE

DISSERTATION SUBMITTED 2020

AALBORG UNIVERSITY
DENMARK

IMMUNOHISTOCHEMICAL DETECTION OF KI67 IN BREAST CARCINOMAS

METHOLOGY, QUALITY ASSESSMENT AND DIGITAL IMAGE ANALYSIS

BY
RASMUS RØGE

AALBORG UNIVERSITY
DENMARK

DISSERTATION SUBMITTED 2020

Author

Rasmus Røge, MD
E-mail: rr@rn.dk

Education

2005-2006: Research year, Department of Nephrology and Clinical Immunology, Aarhus University Hospital, Skejby, Denmark
2009: Cand. Med., Aarhus University, Denmark
2013-2020: Ph.d.-student, Aalborg University, Denmark

Employment

2007-2009: Department of Oncology, Aalborg University Hospital, Denmark
2009-2010: Department of Infectious Diseases, Aalborg University Hospital, Denmark
2010: Psychiatric Hospital, Aalborg University Hospital, Denmark
2011: Department of Pathology, Aalborg University Hospital, Denmark
2011: Psychiatric Hospital, Aalborg University Hospital, Denmark
2012-2017: Department of Pathology, Aalborg University Hospital, Denmark
2017-2018: Department of Pathology, Regionshospitalet Viborg, Denmark
2018-: Department of Pathology, Aalborg University Hospital, Denmark
2012-: Scheme organizer, NordiQC, Department of Pathology, Aalborg University Hospital, Denmark

English summary

Breast cancer is the most common cancer in women. Mortality have declining in the recent decades due to advances in treatment and diagnosis.

Ki67 is a nuclear protein expressed in proliferating cells, where it acts as a surfactant that prevents chromosomal collapse during mitosis. This unique cyclic expression pattern enables identification of proliferating cells in histological slides by immunohistochemical (IHC) staining for Ki67. The Ki67 Proliferation Index (PI) (fraction of Ki67 positive cells) is an important prognostic marker in breast carcinomas, where increased Ki67 PI is associated with higher mortality and risk of relapse. Ki67 PI have also been suggested as surrogate marker to distinguish between two molecular subtypes of breast carcinomas: Luminal A and B.

However, correct and reproducible assessment of Ki67 PI is challenging due to significant variance in IHC staining procedures and Ki67 interpretation practices among pathology laboratories. The general aim of the studies included in this thesis was to address some of these standardization issues.

In study 1, we examined the interobserver variation of Ki67 PI scoring in a large cohort (n=199) of pathology laboratories that examined the same fifteen Ki67 IHC stained breast carcinomas. Overall, there was good agreement between the participants, but when applying a cutoff of 20% only modest kappa values were observed. Participants that estimated Ki67 PI in hotspots reported higher values than participants that scored an overall average. Similarly, participants who considered weak Ki-67 nuclear staining positive obtained higher Ki-67 PIs than those who did not. A computer algorithm (Virtual Double Staining (VDS)) obtained Ki-67 values similar to the mean value of participants.

In study 2, we compared a Digital Image Analysis (DIA) algorithm VDS that fuses two parallel sections (one for tumour identification and one for Ki67 PI estimation) with stereologically obtained Ki67 PI by a human observer. 140 breast carcinomas were examined, and we found almost perfect correlation between the two methods (ICC > 0.97).

In study 3, we examined the impact of selection of primary antibody clones, antibody formats and stainer platforms on Ki67 PI obtained by VDS in breast carcinomas. Significant differences in Ki67 PIs were observed between the different antibody, format and platform combinations. Similarly, variations in proportion of carcinomas with high level Ki67 (above 20%) were seen indicating that IHC methodology can be an important source of variation.

In study 4, we examined the potential of cell lines as supplementary Ki67 IHC performance controls. Ki67 IHC was assessed as H-scores by two different DIA programs. We found low day-to-day variation in Ki67 IHC H-scores in normal slides. Suboptimally stained slides could be identified on the basis of lower H-scores.

Dansk resumé

Brystkræft er den mest almindelige kræftform blandt kvinder. Dødeligheden er faldet i de seneste årtier på grund af fremskridt inden for diagnosticering og behandling.

Ki67 er et kerneprotein, der udtrykkes i delende celler. Her fungerer det som surfaktant, der forhindrer kromosomalt kollaps under mitosen. Dette unikke cykliske udtryksmønster muliggør identifikation af prolifererende celler i immunohistokemisk (IHC) Ki67-farvede histologiske snit. Ki67 Proliferation Index (PI) (fraktionen af Ki67-positive celler) er en vigtig prognostisk markør i mammakarcinomer, hvor øget Ki67 PI er forbundet med højere dødelighed og risiko for tilbagefald. Ki67 PI har desuden været foreslået som surrogatmarkør til at skelne mellem to molekylære subtyper af mammakarcinomer: Luminal A og B.

Korrekt og reproducerbar vurdering af Ki67 PI er imidlertid udfordrende på grund af betydelig variation i IHC-procedurer og Ki67-fortolkningspraksis blandt patologilaboratorier grundet manglende standardisering. Det overordnede formål med studierne inkluderet i denne afhandling var at undersøge nogle af disse standardiseringsudfordringer.

I studie 1 undersøgte vi interobservatør variationen af Ki67 PI i en stor kohorte (n=199) af patologilaboratorier, der undersøgte de samme femten digitaliserede Ki67 IHC-farvede mammakarcinomer. Generelt var der en god overensstemmelse mellem deltagernes resultater, men ved anvendelse af et cutoff på 20% observerede vi kun beskedne kappa-værdier. Deltagere, der estimerede Ki67 PI i hotspots, rapporterede højere værdier end deltagere, der scorede et samlet gennemsnit. Tilsvarende opnåede deltagere, der betragtede svage Ki67 kernefarvninger som positive, højere Ki-67-PI end dem, der opfattede dem som negative. Computeralgoritmen Virtual Double Staining (VDS) opnåede Ki-67 PI svarende til deltagernes gennemsnit.

I studie 2 sammenlignede vi en Digital Image Analyse (DIA) algoritme, VDS, der fusionerer to parallelle snit (et til tumoridentifikation og et til Ki67 PI-estimering) med stereologiske estimater af Ki67 PI fra en menneskelig observatør. 140

mammakarcinomer blev undersøgt, og vi fandt en næsten perfekt overensstemmelse mellem de to metoder (ICC > 0,97).

I studie 3 undersøgte vi effekten af forskellige primære antistofkloner, antistof-formater og farveplatforme på Ki67 PI i mammakarcinomer, der blev estimeret ved hjælp af algoritmen VDS. Vi observerede signifikante forskelle blandt de forskellige antistofkloner, formater og platformkombinationer. Tilsvarende observerede vi betydelige variationer i andelen af karcinomer med højt niveau af Ki67 PI (over 20%), hvilket indikerer, at IHC-metodologien er en vigtig kilde til variation.

Endeligt undersøgte vi i studie 4 potentialet for cellelinjer som Ki67 IHC performancekontroller. Ki67 IHC blev vurderet ved hjælp af en H-score beregnet af to forskellige DIA-programmer. Vi fandt lav dag-til-dag variation i Ki67 IHC H-score. Suboptimalt farvede snit kunne desuden identificeres på basis af lavere H-score.

Acknowledgements

List of abbreviations

Ab   Antibody
DIA   Digital Image Analysis
IHC   Immunohistochemistry
PCK   Pan Cytokeratin
PI (PIs)   Proliferation Index (indices)
ROI   Region of interest
VDS   Virtual Double Staining

Content

Table of figures

1 Scientific papers

This PhD thesis is based on four independent scientific papers and referred to in thesis using Arabic numbers (1-4) in the order listed below.

1.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Ki67 proliferation index in breast cancer as a function of assessment method. A NordiQC experience. Appl Immunohistochem Mol Morphol. 2020;10.1097/PAI.0000000000000846.
doi:10.1097/PAI.0000000000000846. Online ahead of print.

2.
Røge R, Riber-Hansen R, Nielsen S, Vyberg M. Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining. Breast Cancer Res Treat. 2016 Jul;158(1):11-19. doi: 10.1007/s10549-016-3852-6.

3.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Impact of Primary Antibody Clone, Format, and Stainer Platform on Ki67 Proliferation Indices in Breast Carcinomas. Appl Immunohistochem Mol Morphol. 2019 Nov/Dec;27(10):732-739.

4.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Potential of Cell Lines Assessed by Image Analysis Used as Performance Controls of Ki67 Immunostained. Appl Immunohistochem Mol Morphol. 2020 Mar 9. doi: 10.1097/PAI.0000000000000845. Online ahead of print.

2 Background

2.1 Ki67

2.1.1 Gene and structure

Ki67 is a nuclear protein encoded by the MKI67 gene located on the long arm of chromosome 10 [1]. The name originated from a laboratory in Kiel that produced the prototype antibody (Ab) in well 67 of a 96-well plate using a HL-60 (Hodgkin lymphoma) cell line [2]. In humans, it exists in two isoforms with the molecular weights of 320 kDa and 359 kDa [3].
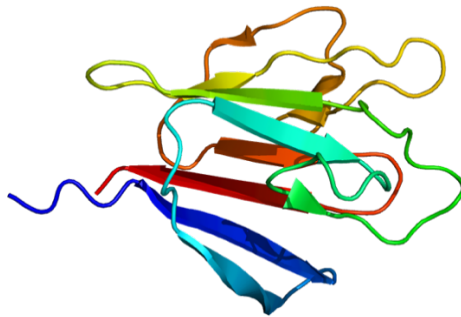
Figure 2.1. Ultrastructure of Ki67 protein. (courtesy of Emw - released under creative commons 3.0)

The MKI67 gene is found in all species of vertebrates [4]. It consist of several different regions [5]: N-terminal forkhead associated (FHA) domain, phosphatase binding (PB) domain, tandem repeats containing domain and a leucine and arginine-rich (LR) C-terminal. The FHA is a phosphopeptide recognition domain, which is found in many regulatory proteins that generally are involved in assembly of protein complexes [6,7]. The phosphatase binding domain interacts with certain isoforms of protein phosphatase 1, which, besides many other functions, protect the centromeric cohesion from phosphorylation during mitosis [8,9]. The central region of Ki67 consists of 16 tandem repeats that are phosphorylated during mitosis by cyclin dependent kinase 1 [5,10]. This makes the protein highly positively electrostatic charged. The LR domain contains many leucine and arginine amino acids that have the ability to bind to DNA [11]. This domain also binds to heterochromatin 1 that are

involved in transcription, DNA repair and nuclear architecture [12,13].

Figure 2.2. Ki67 gene structure. (FHA: Fork head associated domain, PB: phosphatase binding domain, LR: Leucine/arginine rich domain)

2.1.2 Function and expression pattern

The function of Ki67 eluded the scientific community for many years. From the localization of the protein and the cyclic phosphorylation it was clear that it was related to cell division. However, the exact nature behind especially the tandem repeats remained unclear until a landmark publication in 2016, where Cuylen et al. showed that Ki67 acted as a chromosomal surfactant preventing chromosomes from collapsing after disassembly [14], which is crucial for chromosome motility during cell division. The structural explanation behind this is the pattern of Ki67 interaction with DNA. The positively charged repeated tandems within the central part of Ki67 cover the surface of the chromosomes and will, due to the positive charge, repel each other. During mitosis Ki67 comprise the majority of this perichromosomal layer formed together with other proteins [15]. The LR region binds to DNA, while the FHA and PB regions interact with regulatory proteins in the nucleoplasm.
Recently, it has also been shown that Ki67 play a role in regulating cell proliferation during interphase [16].

Sobecki et al. showed that Ki67 protein expression increases throughout the G2-phase and peaks in M-phase of the cell cycle [17] in both normal and neoplastic cells. This was also reflected in the mRNA expression, which peaked in G2. During the G1-phase Ki67 is degraded although some Ki67 remain until cells enter full quiescence. During the interphase, Ki67 is located to nucleoli [16,18]. This pattern is also observed in IHC staining slides where some cells (most likely cells in M- or early G-phase) are completely stained, while other cells only have a dot-like reaction pattern due to accumulation in the nucleoli.

2.2 Immunohistochemistry

The discovery of Ki67, based on the anti-Ki67 (Ki-67) Ab, was a landmark in the immunohistochemical (IHC) field. It was now possible to distinguish between proliferating and non-proliferating cells. A number of studies correlated proliferation and outcome and found prognostic values of Ki67 [19,20]. As an example, Hall et al. found worse survival in patients with a low-grade Hodgkin lymphoma with a high percentage of Ki67 positive cells [19]. Unfortunately, the prototypic Ab was limited to fresh and frozen tissue only.

Since then many new antibodies (both polyclonal and monoclonal) have been developed, which can also be used on formalin-fixed tissue. The most commonly used Ab is Mib1 that was introduced alongside Mib2 and Mib3 by Cattoretti et al. [21]. The IHC staining protocol did now include a Heat-Induced-Epitope-Retrieval (HIER) step, where the tissues were boiled in low pH buffers [22]. According to NordiQC that examined the IHC staining quality of Ki67 in 2016, 54% (220/409) of the participants used this clone for detection of Ki67 [23]. Other commonly used commercially available clones are MM1 (Leica/Novocastra, UK), SP6 (rabbit monoclonal, available from many vendors) and 30.9 (rabbit monoclonal, Roche/Ventana, Tucson, US). The vast majority of staining protocols included epitope retrieval. None of the participants used the prototypic Ab. The antibodies are available both as pre-diluted Ready-To-Use and concentrated formats. In the most recent NordiQC Ki67 assessment, 39.6% (162/409) of participants used a concentrated format of a Ki67 clone, compared to 60.4% that used an RTU product.

Concerning controls, the international ad hoc expert committee on IHC at the moment recommends tonsil as IHC critical assay performance controls (iCAPC), where all germinal center B-cells and basal squamous epithelial cells must show a moderate to strong nuclear staining [24]. As supplementary tissues, appendix (staining reaction should be seen in basal crypt cells) and liver (negative control) are recommended. The study also underlines that the positive controls for Ki67 should include cells/tissues with the relevant clinical level of Ki67 expression [24].

2.3 Breast cancer

According to the latest WHO World Cancer Report (2014), breast cancer is the most common cancer among women [25]. The incidence varies between 70 and 90 cases per 100,000 women in

industrialized countries. It was estimated that the number of new cases in 2012 was 1.7 million, and that breast cancer constituted 25% of all cancers in women. 15% of cancer related deaths was attributed to breast cancer. In newly industrialized nations, both incidence and mortality are rising. The case fatality is highest in less developed countries. However, the overall mortality rates have been declining since the early 1990, which is attributed to earlier detection and more effective treatment regimes.

For Denmark the incidence of breast cancer during 2012-16 was 92.3, i.e., the average number of new cases was 4,700 (24.3% of all new cancers) [26]. The age standardized death rate (2012-2016) was 16.3%. During the latest 10 years with available numbers the death rate was reduced by 3.4%. The 1- and 5-year relative survival was 97% and 87%, respectively.

The etiology of breast cancer is multifactorial, mostly related to endocrine hormonal factors related to adiposity, first delivery after age of 30, nulliparity, use of some contraceptives or menopausal hormone replacements [25]. A minor fraction of breast cancer cases can also be associated with genetic predispositions, namely mutations in the BRCA1 and BRCA2 genes. Other gene mutations inducing breast cancer with lower penetrance have also been identified.

The term breast cancer covers a heterogenous spectrum of neoplasms both in terms of clinical behavior and histomorphological appearance. The diagnosis of breast cancer is most often made by an experienced pathologist on clinical samples such as biopsy, lumpectomy or mastectomy material. The vast majority of breast neoplasms are epithelial in origin. Often in situ lesions are seen alongside invasive breast cancer.

There exists many subtypes of epithelial breast carcinomas – the latest WHO classification of breast tumours (2014) lists more than 25 subtypes of carcinomas [27]. However, by far the most common type is the invasive ductal carcinoma, NST (no special type), which makes up 70-80% of breast neoplasms. This non-special subtype does not contain the specific morphological characteristics required for the other special subtypes and is the default subtype but is morphological heterogenous. The most common special subtype is the lobular carcinoma accounting for around 15% of invasive breast neoplasms [28]. In general, lobular carcinomas have a good prognosis due to low histological grade and expression of hormone receptors (see later). However, some studies indicate that high grade lobular carcinomas have worse prognosis compared to ductal

carcinomas, NST of the same grade [28]. Additionally, lobular carcinomas can be difficult to detect in screening due to a diffuse pattern of invasion.

Breast carcinomas (of NST and lobular types) are graded based on the Nottingham grading system that assesses three histomorphological parameters of the tumour: 1) the degree of architectural differentiation (formation of tubules), 2) nuclear pleomorphism and 3) number of mitosis [29,30]. This grading provides prognostic information with worse prognosis for high grade tumours [31].

Operable breast cancers are treated with either mastectomy or lumpectomy [32]. Systemic treatment after operation are offered to patients depending on the results of the pathological examination of the primary tumour and lymph nodes. In general, patients with Estrogen receptor (ER) positive tumours are treated with endocrine therapy except patients with very low risk of recurrence (small tumour, low Nottingham grade, no lymph node metastases, human epithelial growth factor receptor 2 (HER2) negative) [33]. Patients with HER2 positive carcinoma are treated with chemotherapy and trastuzumab. Chemotherapy are offered for patients with ER-positive tumours and certain risk factors and/or molecular subtypes (see 2.5).

2.4 Biomarkers in breast cancer

Breast carcinomas are routinely examined for a number of biomarkers. The earliest identified was ER that is now considered the most important prognostic and predictive biomarker in breast cancer. Approximately, 75-80% of breast carcinomas are IHC-positive for ER. ER-positive breast cancers are shown to have better survival than negative cancers [34,35]. In the early eighties, the first study was published on adjuvant tamoxifen (selective estrogen receptor modulator) therapy that showed improved survival in breast cancer [36]. Since then, it has been shown many times that ER-positive tumours benefit from endocrine therapy [37].

Many countries (but not Denmark) also test breast carcinomas for progesterone receptor (PR). Since PR is a downstream effector of ER signaling, PR only adds little predictive information [38]. A few studies indicate increased benefit of endocrine therapy in ER+/PR+ positive breast cancers [39]. The controversial ER-/PR+ entity may

benefit from endocrine therapy [40], although the existence of this subtype has been questioned [41].

HER2 protein overexpression (as detected by IHC) and gene amplification (as detected by in situ hybridisation) was seen in 15-20% of breast carcinomas in an unselected Finnish population [42]. It is a prognostic marker of aggressive course of disease [43]. Since the development of trastuzumab (anti-her2 Ab), HER2 has also become a predictive biomarker of response to treatment [44,45].

Ki67 is an important prognostic marker in breast cancer. In a meta-analysis of 46 studies with 12,155 patients with early breast cancer, the authors found significantly higher overall risk of relapse (HR = 1.93, CI: 1.74-2.14) in patients with Ki67 positive cancers [46]. This was also seen when stratifying for lymph node status (negative, HR = 2.31; positive, HR = 1.59). Similar results were also seen for survival. The cutoff between Ki67 positive and negative was defined by the authors of the included studies. Stuart-Harris et al. published a review on proliferation markers in early breast cancer and found that high Ki67 had an impact on overall and disease-free survival [47].

Recently, a meta-analysis of 35 studies of triple negative breast carcinomas (i.e., negative for ER, HER2 and PR) found that high Ki67 expression (cutoffs varying between 10-50%) was associated with lower disease-free survival (HR = 1.73, CI: 1.45-2.07) [48].

2.5 Molecular subtypes in breast cancer

In the beginning of the new millennium, a few landmark articles was published that identified several molecular subtypes of breast cancer based on hierarchical clustering of gene expression [49–51]. A new finding was that the ER-positive cancers could be subdivided into subgroups with different prognosis. They described so-called intrinsic subtypes were named Luminal A, Luminal B, HER2-enriched and Basal-like. There was significantly differently survival among the different groups. The HER2-enriched and Basal-like have an aggressive course but are responsive to chemotherapy [38]. The Luminal A have more favorable prognosis but is resistant to chemotherapy, while Luminal B are aggressive and have varying chemo responsivity [38]. Extensive genomic analysis have validated the four major molecular subgroups in regard to both genetic and epigenetic abnormalities [52].

In a study by Cheang et al., the classical IHC markers ER, PR, HER2 and Ki67 appeared to distinguish breast carcinomas Luminal type A (ER+, PR+/-, HER2-) and B (ER+, PR+/-, HER2+/-) when compared to gene expression profiles in a test cohort [53]. The best Ki67 cutoff between the two groups were 13.25%. In an independent cohort, Luminal A defined on the IHC classification had significantly better survival than Luminal B (and Luminal HER2+) [53]. Another study found no benefit from adjuvant chemotherapy in Luminal A breast cancers in cohort of high risk premenopausal patients randomized to either adjuvant or no treatment [54]. On the other hand, Luminal type B benefitted from treatment.

Several commercially available genomic assays have since become available that based on expression of genes estimate risks of recurrence: Oncotype DX is a reverse transcriptase PCR assay that examines the expression of 16 genes related to ER, HER2, proliferation and invasion (and reference genes) [55]. The Oncotype DX produces a recurrence score between 0-100, where higher scores indicate a higher probability of recurrence. PAM50 is based on Nanostring technology with probes that bind directly to mRNA. The gene expression of 50 genes are measured and a Risk of Recurrence score and the intrinsic subtypes is reported [56].
Others genomic assays mentioned in the literature are MammaPrint, EndoPredict and Breast Cancer Index.

These assays are useful for identifying patients with lower risk of recurrence which could indicate that no adjuvant chemotherapy is needed.

2.6 Ki67 as a surrogate marker of Luminal A

Since molecular testing is not available in many clinical settings, pathologist and oncologists for the time need to rely on the surrogate IHC markers for selection of treatment regimes. It is often possible, based on IHC markers, to identify some of the patients that should be allocated to chemotherapy treatment, especially the triple negative and HER2-positive tumours (ER +/-) [57]. Among patients with ER-positive tumours, chemotherapy treatment decision is based on risk factors (low age, lymph node metastases etc.), where lymph node status seems to be the most predominant determinant [58].

The St. Gallen international expert consensus group in 2011 introduced a suggested Ki67 cutoff of 14% for discrimination between luminal type A and B based on the previous mentioned study with comparisons of IHC profile with the PAM50 intrinsic subtype [53,59]. This was in the later St. Gallen recommendations from 2013 changed to "Ki67 low" for Luminal type A and "Ki67 High" for Luminal type B in recognition on the many clinically challenges in documenting this cutoff [60]. In 2015, a majority of panelists at St. Gallen was prepared to accept a Ki67 cutoff in 20-29% range, when taking in account knowledge on the local laboratory values [57]. A majority of panelists did not believe that multiparameter molecular markers were required for this distinction [57]. The criteria concerning Ki67 did not change in 2017 [61] although Luminal type B could also be classified based on high histomorphological grade and a "bad" molecular marker. Additionally, an intermediary group between luminal type A and B were introduced. Most recently in 2019, the St. Gallen panelist strongly endorsed the value of genomic assays in determining whether to recommend chemotherapy [62].

However, the Ki67 has been challenging to score both in regard to separation of positive and negative nuclei but also where to score. Scoring reproducibility varies significantly [63] and conflicting studies have been published concerning hotspots versus overall average [64,65]. For this reason, alternative and standardized methods for obtaining Ki67 proliferation indices (PIs) are attractive.

2.7 Digital Image Analysis

2.7.1 Fundamentals Digital Image Analysis

Digital images can principally be separated in two fundamental different types: vector- and raster-images. Vector-images (e.g. .pdf-files) are composed of geometric shapes with information on individual colors and have the advances of being infinitely scalable and relatively small in size. On the other hand, raster-images (e.g. .jpeg and .tiff-files) can be considered as a large chessboard with stored color information in each field (pixel). For greyscale images (e.g. radiology images) one value is stored while several values are stored for each pixel in color images depending on the color model used. This makes stored digital images large in data size and with limited possibilities of zooming. Raster-images are the principal format recorded from digital cameras and slide scanners using in pathology.

Digital Image Analysis (DIA) is the process of extracting information from raster images by analyzing the pattern of color information and spatial location of the pixels in a reproducible manner [66]. The output from DIA can take many forms: Simple measurements or counting of pixels, identification of regions with specific characteristics or complex quantification analyses of textures and shapes [66].

This DIA process can theoretically be split up in four analysis steps: preprocessing, segmentation, postprocessing and output of calculations [67].

The purpose of the preprocessing step is to optimize the image for remaining analyses and reduce the complexity of the image. This is accomplished by different mathematical filters that remove irrelevant colors or noise (e.g. dust) in the image. It is also possible to enhance certain structures (e.g. edges) [67]. Examples of noise and edge-filters can be seen in Figure 2.3.
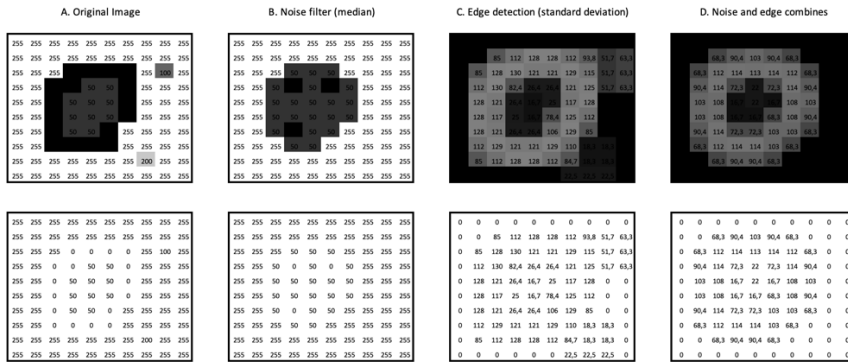
Figure 2.3. Image analysis algorithms applied on a grey scale image. Bottom row: Grey scale values. Top row: Visual interpretation. A. Original Image. B. Noise filter (median) applied on original image. C. Edge detection filter (standard deviation) applied on original Image. D. Both filters applied on original image.

During the <u>segmentation</u> step, pixels of the image are labeled into groups based on common characteristics such as values in the color filters or spatial relation to other pixels (indirectly a measurement morphological characteristics) [67]. Many different segmentations algorithm can be used. Simple thresholds of color intensity values are often used to separate pixels into categories (e.g. IHC positive, IHC negative or background). However, thresholds can be sensitive to variations in staining intensity and may lose information on nuclei shapes [68]. Another example of segmentation is use of a supervised Bayesian algorithm, where the algorithm is trained based on pixels from different regions of interest (e.g. tumour tissue, normal tissue and background) [69]. The algorithm is a multivariate probabilistic classifier that segments all pixels on the similarities to the different training groups. A third example is the K-means clustering algorithm that unsupervised segments pixels into K clusters based on the Euclidian distances of the intensities in the color space [67].

The <u>postprocessing</u> step alters the results of the segmentation and is used to improve the output. For example, similar pixels areas located closed to each other can be fused or closely situated nuclei detected as one can be separated.

Finally, calculations from the algorithm are output. These are typically in the form of a number (number of positive/negative cells or areas), area (size of different categories of area) or in more advanced algorithms probability of certain diagnosis.

Finally, recent advances in machine Learning have produced numerous segmentation algorithms that reliably are able segment e.g. melanomas from nevus on par with pathologists [70–72]. One drawback to these artificial intelligence algorithms are the large dataset required for accurate training.

2.8 Image analysis for Ki67

With the recent advances in technology, many studies have examined the feasibility of DIA for estimation of Ki67 PIs. One of the first published methods was ImmunoRatio (free version closed 2019) that was an online browser-based web application where users could upload IHC staining images [73]. The application algorithm would segment nuclei as either positive or negative and summary statistics including Ki67 PI. The algorithm could also be used for ER and PR. This method was applied in a study of 577 breast carcinomas, where the authors found excellent correlation between a human observer and ImmunoRatio also in regard to a Ki67 cutoff of 20% (kappa 0.881) [74]. However, comparison with another DIA method (Cell Counter plugin for ImageJ) revealed significant discrepancies in categorization breast cancer (cutoff 14%) [75]. Many commercially available algorithms are also available (e.g. Aperio ImageScope (Leica) and Companion (Ventana)). Studies have also found good correlations between human observers and these algorithms [76,77].

Applications have even been developed for smartphones mounted on microscopes for automatic Ki67 PI estimation which showed excellent correlation with both human observers and ImmunRatio [78].

3 Aims and hypothesis

The general aim of this thesis was to address some of standardization issues that challenge the IHC marker Ki67 that is used for calculating Ki67 PIs, which are important in subclassification of breast cancer. We wanted to examine the applicability of DIA (specifically the VDS algorithm) as a tool for estimation of Ki67 PI in IHC stained slides and compare the results with human observers. Furthermore, we wished to examine the necessity of standardization in IHC staining for Ki67 in tissues stained with different IHC protocols. Finally, we wanted to develop a method for monitoring the diagnostic sensitivity of Ki67 IHC using DIA in standardized materials.

Hypotheses
1) Ki67 PI estimated by human observers vary significantly.
2) Ki67 PI obtained using VDS in IHC stained breast cancers are similar to results from a human observer estimating Ki67 PI using stereological counting.
3) Ki67 PI obtained using VDS in optimally IHC staining breast cancers vary depending of the primary Ab clone, format and stainer platform applied.
4) On slide cell culture samples can be used as IHC sensitivity control.

4 Materials and methods

This PhD project aimed to examine different phases of the Ki67 IHC staining and interpretation process and establish whether new or standardized methods could improve the quality or reliability of the Ki67 IHC staining and/or interpretation.

The first study (1) was a cross sectional study on inter-observer interpretation variation and approaches for estimation of Ki67 PIs in breast carcinomas in a large cohort of clinical pathology laboratories. Data from the participants were compared to the results from the DIA Virtual Double Staining algorithm established in study 2.

The purpose of the second study (2) was to compare Ki67 PIs estimates obtained by DIA, specifically the Virtual Double Staining algorithm, to a human observer. Manual estimates for Ki67 PIs was quantified using the current stereological gold standard (counting frames in systematic uniform random sampled areas).

The third study (3) aimed to examine the impact of IHC protocol parameters on the Ki67 PI in a cohort of breast cancer patient. The included samples were stained for Ki67 (and PCK) using different commercially available anti-Ki67 monoclonal Abs both as concentrated format and ready-to-use on the most commonly used stainer platforms. The Ki67 PIs in the clinical tissues were obtained using the VDS algorithm from study 2.

The final study (4) was a technical experiment aimed to examine the potential of cell lines as supplementary Ki67 IHC performance controls. Ki67 IHC was assessed by calculating H-scores using two different DIA programs. Both day-to-day variation and effect of suboptimal staining protocols were examined.

4.1 Tissue materials

4.1.1 Study 1

Tissue samples from 41 breast carcinomas obtained from the archives of the Department of Pathology at Aalborg University Hospital, Denmark. After initial Ki67 IHC staining, 15 samples were selected from these to include breast carcinomas with even distribution of Ki67 PIs from 0% to 100%. The cases comprised 14 ductal carcinomas and one lobular carcinoma.

4.1.2 Study 2

Tissue samples comprised 140 resection specimens obtained consecutively from the archives of the Department of Pathology at Aalborg University, Denmark. Of the 103 cases that was included in

the final experiment, the majority (n=83) was ductal carcinomas. Also included were 13 lobular carcinomas and seven other subtypes (combined ductal and lobular carcinoma, tubular carcinoma etc.)

4.1.3 Study 3
Tissue samples comprised 41 breast carcinomas obtained from the archives of the Department of Pathology at Aalborg University Hospital, Denmark. The histological subtypes of breast carcinomas comprised 34 ductal carcinomas, five lobular carcinomas and two unclassified carcinomas. The majority of all carcinomas was estrogen receptor positive (90.3%, 80.0% and 100% for the three groups, respectively).

4.1.4 Study 4
Cell lines were purchased from Horizon Discovery (UK) and consisted of three different sedimented cell cultures with different levels of Ki67 expression. The materials had been formalin-fixed and paraffin embedded and was received as TMA cores ready to be inserted in donor blocks. Additionally, to validate the clinical consequences of suboptimal protocols, the experiment also included slides with well characterized tissue from the NordiQC Ki67 run B22 (breast cancer module) [23]. In brief, the clinical tissues were three breast carcinomas (with negative, weak and strong Ki67 expression), liver, pancreas and tonsil.

4.2 Tissue fixation and preparation
In general, clinical material had been processed according to the ASCO/CAP guidelines for ER [79]. All clinical materials were routinely fixed for 24-48 hours in 10% neutral-buffered formalin and subsequently embedded in paraffin.

4.3 Tissue Micro Array (TMA)
Before construction of the TMAs, HE whole-slides of included tissue were screened, and areas of interest containing invasive carcinoma but not carcinoma in situ or normal epithelium were marked. Cores from the original block were drilled out using the automated microarray device from 3DHISTECH (Hungary). The diameter for all cores in study 1-3 was 2 mm. In study 4, TMA cores containing the clinical materials measured 4 mm, while the cell cultures measured 2.5 mm (one core) and 1.75 mm (two cores).
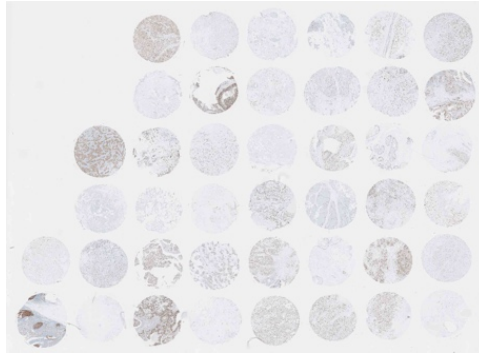
Figure 4.1. Tissue micro array - 41 cores of breast carcinoma and one tonsil.

All drilled out cores were transferred to recipient paraffin blocks and heated in an oven for 20 min. at 37°C to fuse the material. For study 1 and 3, all cores were collected in one TMA. Due to the number of cores, three separate TMAs were made for study 2.

4.4 Immunohistochemistry

All IHC stains were conducted at the Department of Pathology at Aalborg University Hospital, Denmark. All cut slides were 3 micrometer and mounted on SuperFrost+ (Menzel Gläser, Germany).

4.4.1 Study 1

Two neighboring sections were stained for either Ki67 or PCK using the protocol settings presented in Table 4.1.

Epitope	Ki67	PCK
Antibody clone (dilution)	Mib1 (RTU)	AE1/AE3 (cocktail) (1:100)
Vendor	Dako Agilent	Dako Agilent
Stainer platform	Dako Autostainer Link 48	Dako Autostainer Link 48
Epitope retrieval / buffer	HIER, TRS low pH	TRS High pH
Visualisation system	EnVision FLEX	EnVision FLEX+ (mouse)

Table 4.1. Immunohistochemical protocol parameters in Study 1.

4.4.2 Study 2

Two neighboring sections were cut from each TMA and staining for either Ki67 or PCK using the protocol settings presented in Table 4.2. Additionally, 5 neighboring slides were stained using PCK.

Epitope	Ki67	PCK
Antibody clone (dilution)	Mib1 (1:200)	AE1/AE3 (cocktail) (1:100)
Vendor	Dako Agilent	Dako Agilent
Stainer platform	Ventana Benchmark Ultra	Ventana Benchmark Ultra
Epitope retrieval / buffer	HIER, CC1	HIER, CC1
Visualisation system	ultraView DAB	ultraView DAB

Table 4.2. Immunohistochemical protocol parameters in Study 2

4.4.3 Study 3

Commercially available, commonly used anti-Ki67 Abs were examined in this study, see Table 4.3. Some products were available both as prediluted RTU and concentrated formats (Mib1 and MM1), while others (SP6 and 30.9) only were sold as either concentrated or RTU format. For the concentrated formats, optimized staining protocols were developed for each stainer platform (Ventana Benchmark Ultra, Dako Autostainer 48 and Leica Bond) based on the recommendations from the international IHC ad hoc expert panel [80]. A single optimized PCK protocol was developed for each platform. During the optimization process, the concentrated format for the clone MM1 displayed aberrant background staining. No replacement was available from the vendor, and this product was excluded from further analysis. For each Ab clone, format and stainer platform combination as set of neighboring slides was staining for Ki67 and PCK.

Primary antibody clone	Mib1	Mib1	SP6	30.9	MM1
Format	Concentrated	RTU	Concentrated	RTU	RTU
Venta Benchmark Ultra	+		+	+	
Dako Autostainer 48	+	+	+		
Leica Bond	+		+		+

Table 4.3. Immunohistochemical staining in Study 3 (RTU: Ready-to-use)

4.4.4 Study 4

In order to examine the day-to-day IHC Ki67 staining variation, three sets of ten slides were stained on three separate days in the same ten slide trays on the Ventana Benchmark instrument. The

slides were stained using the laboratory standard Ki67 protocol based on Mib1 (diluted 1:200, HIER in CC1 and OptiView (Ventana) as visualization system). Additionally, in order to emulate suboptimal protocols, five sets of two slides were stained using the same protocol but with the Ab diluted from 1:100 to 1:1600.
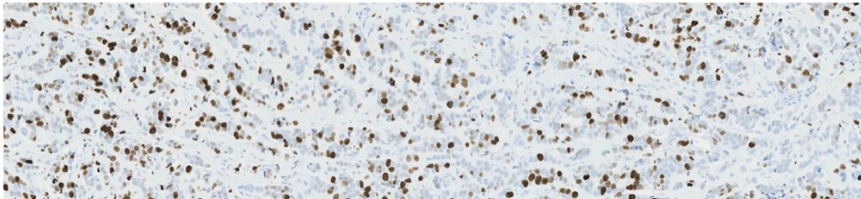
Figure 4.2. Immunohistochemical staining for Ki67 of a breast carcinoma (mAb clone SP6).

4.5 Digitalization of stained slides

All slides were digitized using a NanoZoomer XR (Hamamatsu Photonics, Japan) at the highest resolution (40x). Regions containing tissue were manually selected, while the scanner automatically identified a single focus plane.

4.6 Digital Image Analysis

4.6.1 Virtual Double Staining

The VDS algorithm was developed for IHC quantification of different breast cancer related epitopes (HER2, ER, PR and Ki67). The purpose of the VDS is to limit analysis to tumour cells only. Similar to many physical double stains, the VDS utilizes one IHC staining reaction for identification of cells of interest and a second for quantification. However, the algorithm requires two separate IHC-stained slides. In the studies included in this thesis, slide pairs were stained for PCK (for detection of tumour areas) and Ki67 (quantification).

The VDS algorithm consists of three fundamental steps: First, the algorithm aligns the two slides so that tissue structures will be as parallel as possible. This is obtained using rotation and local deformation of the image. Secondly, tumour areas are detected in the PCK stained slide. This part of the algorithm is based on a Bayesian classifier with stored training information. Finally, cell nuclei (when applied for Ki67) are detected and segmented to either IHC positive or IHC negative. The algorithm could also split up the positive cells into three categories (based on the intensity / "brownness" of the IHC reaction).

Two of the authors (MV and RR) from the studies in this Ph.D. study participated in the development of the algorithm – specifically by submitting images manually marked with Ki67 positive or negative cells and by reviewing initial results from the algorithm. After this initial phase, the algorithm was locked (and commercially available). Hence, the exact Image Analysis steps used for this cell segmentation are not available. However, the developer has stated segmentation of nuclei is based on threshold of the mean intensity of pixels detected in a nucleus.
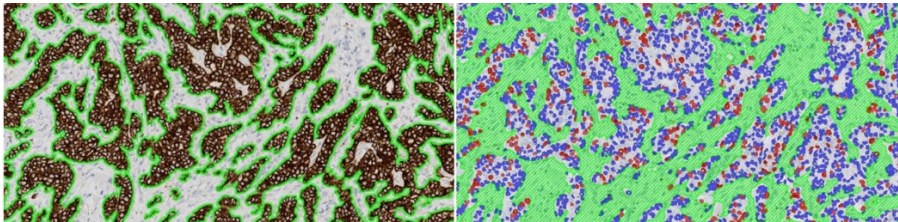
Figure 4.3. Virtual Double Staining principle. Left: Slide stained for Pan Cytokeratin (PCK). Regions of interest (PCK-positive) are detected and transferred to second slide. Right: Ki67 stained slide. PCK-negative areas are ignored (green), while Ki67 positive (red) and negative (blue) are counted.

The complete VDS algorithm was applied on breast cancer tissue in study 1-3. Additionally, both Ki67 PIs and H-scores were calculated in study 3. The main focus in study 4 was Ki67 expression and since the cores only contained cells of interest, image analysis using VIS was limited to the Ki67 segmentation and H-score step.

In order to examine the proficiency of VDS alignment of slides and importance of spatial distance between slides, five neighboring PCK-stained slides were virtually aligned using to VDS algorithm (study 2). An adapted version of PCK algorithm compared both neighbor (1-2, 2-3, 3-4 and 4-5) and non-neighbor slide pairs (1-3, 1-4 and 1-5), and segmented the fused image as "PCK-positive in both slides (+/+)", "PCK-negative in both slides (-/-)" or PCK-positive in one slide (+/-) / (-/+)".

4.6.2   QuPath – study 4

A second DIA program was included in study 4 to examine the consistency in both the estimated Ki67 PIs and H-scores between different algorithms. The program QuPath (University of Edinburgh) was selected since it was freely available, had open source code and was specifically developed for pathological images [81].

The program has an inbuild cell detection algorithm that can segment nuclei in IHC stained slides and provide a cell count, fraction of positive cells (Ki67 PIs) and H-score based on three categories of positivity. All algorithm parameters were the default of the program, except for "Detect Image" where "Optimal density sum", since this parameter qualitatively gave better separation of individual nuclei.

4.7 Manual scoring

Unless otherwise indicated, Ki67 PIs were calculated as the fraction of positive tumour cells ($n_{Ki67\ positive\ tumour\ cells} / n_{tumour\ cells}$) in all studies. A Ki67 positive tumour cell was qualitatively defined as a cell nucleus with an unequivocal IHC staining reaction regardless of size or staining intensity.

4.7.1 Online scoring module – Study 1

In order to examine interobserver variability in K67 PIs estimation, a survey module was created online as part of the run B18. The module was hosted by PathXL (now purchased by Philips and renamed Philips Tutor) and included 15 scanned tissue cores of breast carcinoma. All participants in the NordiQC breast cancer module (n=417) were individually invited to participate. Apart from scoring Ki67 PIs, participants were asked on their method of Ki67 PIs estimation (counting vs. eye balling, hot spot vs. average), job title, clinical control tissue and whether they considered weakly and moderately Ki67 stained nuclei as Ki67 positive.

The participants had one month to complete the challenge. After completion of the module, participants received individualized feedback on their scoring compared to the overall results.

4.7.2 Manual stereological scoring – Study 2

All breast carcinoma cores were manually assessed for Ki67 PIs. Due to the large number of tumour cells in every core, a number of areas were sampled according the Systematic Uniform Randomized Sampling-principle. This step was performed automatically in the VIS-program (Visiopharm A/S, Denmark) using the newCast function. The computer randomly placed a grid of counting frames on each of the cores. All tumour cells within the counting frames was marked as either Ki67 positive or Ki67 negative by a specially trained pathologist (RR). Specifically, cell profiles within the counting frame or touching the upper or right border were included, while cells touching the lower or left border were excluded in order to prevent overestimation of cell profile numbers [82]. Complicated cases were discussed with a senior colleague (MV). The counted

number of positive and negative cells could be extrapolated to a Ki67 PI estimate of the whole core.
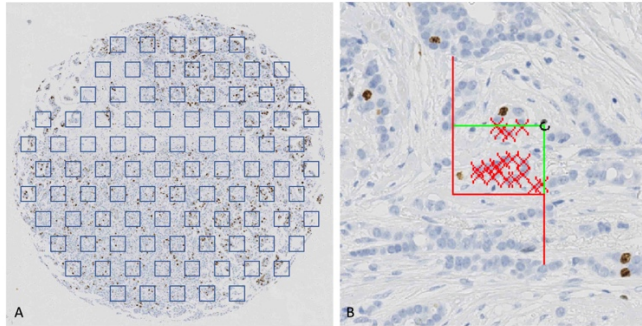
*Figure 4.4. **A**. Systematic Uniform Random Sampling (SURS). Grid randomly placed on a breast cancer tissue core. **B**. Counting frame on a Ki67 IHC stained slide. Each cell within the frame were manually marked as Ki67 positive or negative.*

4.7.3 Quality scores – study 4

The staining quality was assessed in the clinical tissues incorporated in the TMA in study for (see 4.1.4 for composition). An adapted NordiQC scale was applied with the marks: Optimal, Good and Insufficient [83]. Slides were marked according to adherence to the criteria set for Ki67 stained slides in NordiQC Ki67 run B22 [23].

4.8 Statistical analyses

Data preparation and statistical analyses were performed in R (various versions) with the graphical user interface RStudio (various versions) installed. All data formatting and calculations were performed using reproducible scripts (available on request) [84]. Figures were created using the 'ggplot2' package (implementation of Grammar of Graphics) [85]. In boxplots, boxes represent the inter quartile range (IQR), central lines represent median and whiskers represent the most extreme data points within 1.5 IQR [86]. Data points outside this range were considered outliers and represented by dots. In some plot, outliers were removed from plot, but not underlying statistical calculations, to enhance the visual interpretation. P-values lower than 0.05 were considered statistically significant. H-scores were calculated using the formula:

$$H = F_{weak} + F_{moderate} \times 2 + F_{strong} \times 3, \text{ where } F = \text{fraction of cells}$$

The following subsections will briefly introduce statistical methods only used in some of the studies.

4.8.1 Study 1

Ki67 PIs between two or more subgroups were compared using either Mann Whitney U-test or Kruskal-Wallis test. Multivariate analysis was performed using factorial ANOVA. Level of agreement between groups when separating in Ki67 Low or Ki67 High was examined using Fleiss' kappa. Intra Class Coefficients (ICC) for estimation of agreement between the two methods were calculated using the 'psych'-package for R.

4.8.2 Study 2

R2 was calculated using the 'lm' function in R. Bland-Altman plots visualize the mean value of the methods (x-axis) versus the difference between the two methods (y-axis) [87]. ICC was calculated as in study 1 (see 4.8.1).

4.8.3 Study 3

Comparisons of Ki67 PIs and H-scores between groups were accomplished using either ANOVA (with post hoc test: Tukey HSD) or Kruskal-Wallis test (post hoc test: Dunn). ICC was calculated as in study 1 (see 4.8.1).

5 Results

In the sections below, I will briefly summarize the main results presented in the individual studies.

5.1 Study 1

199 individual participants completed the Ki67 scoring webmodule. Data from 77 cores (2.5%) had not been filled in. This was not related to any specific core. In general, there was a moderate to high level of agreement in Ki67 PIs between the observers with an ICC of 0.85.
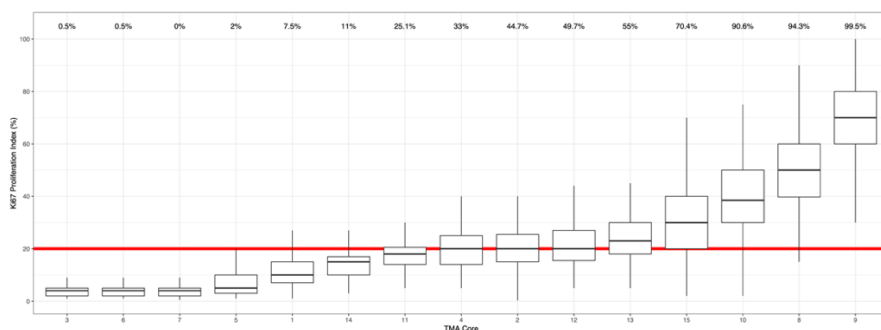
Figure 5.1. Ki67 scoring by 199 participants. Box plots show the variation of scoring. Red line indicates the suggested cutoff 20%. Numbers above each box plots show the percentages of participant scoring a core as Ki67 High (=> 20%). (Adopted from Røge et al. Ki67 Proliferation Index in Breast Cancer as a Function of Assessment Method. A NordiQC Experience. Appl Immunhistochem Mol Morphol (Study 1)).

However, six cores (2,4,11-13 and 15) were challenging for the participant. In these, more than 20% of the participants categorized the tumour differently from the majority. The highest level of disagreement was seen in core 12, where 50.3% categorized this as Ki67 Low. All cores with a high level of disagreement had mean Ki67 PIs close to the cutoff (20%).

There was statistical difference in K67 PIs between observer that were senior pathologists compared to residents in pathology (see Figure 5.2). Observers that scored Ki67 PIs in "hot spot" had higher Ki67 PIs compared participants estimated the overall average. Likewise, observers that considered weakly Ki67 stained nuclei as positive provided higher Ki67 PIs than those who did not. Manual counting (as opposed to eyeballing) provided statistically significant

higher Ki67 PIs (see Figure 5.2). However, there was no significant difference in Ki67 PIs depending on the number of cells counted (100, 200 and 500 cells). When counting more than 1,000 cells, Ki67 PIs tended to be lower, although this was not statistically significant (for most cores).
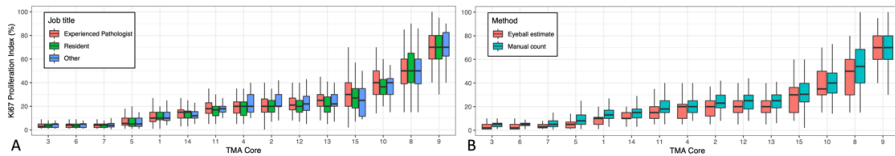
Figure 5.2. Comparison of Ki67 PIs estimates stratified by: A. Job title, B. Method for Ki67 PI estimation. (Adopted from Røge et al. Ki67 Proliferation Index in Breast Cancer as a Function of Assessment Method. A NordiQC Experience. Appl Immunhistochem Mol Morphol (Study 1)).

The results of the participants that estimated the overall average were compared to data obtained from the VDS algorithm. There was a deviation from the average of the participants and VDS between 0.1 and 5.8%. When applying the 20% cutoff, VDS categorized all carcinomas except two (2 and 11) as the majority of participants.
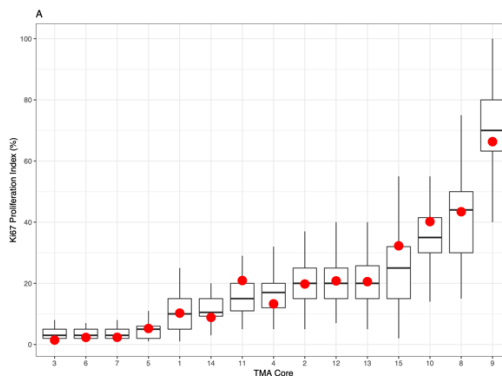
Figure 5.3. Ki67 PIs estimated in 15 breast carcinomas by participants (scoring the overall average of the core) compared to results from Virtual Double Staining (Red dots). (Adopted from Røge et al. Ki67 Proliferation Index in Breast Cancer as a Function of Assessment Method. A NordiQC Experience. Appl Immunhistochem Mol Morphol (Study 1)).

The participants reported that they most commonly (48.3%) used some form of lymphoid tissue (tonsil, lymph node or appendix) as

external IHC control. A large group (33.7%) did not use external controls at all.

5.2 Study 2

Of the 140 breast carcinoma cores included in this study 103 was analyzed further. 37 cores were excluded due to missing tissue, small tumour areas or folding. The VDS algorithm was applied both on the areas sampled for manual Ki67 PI estimation and the whole core. In the sampled areas, the VDS on average identified 14.1 % fewer cells than the human observer.

Overall, there was a good correlation between the human observer, counting using stereological counting frames, and the VDS algorithm (see Figure 5.4). The R^2 for linear fit was 0.96. All relevant ICC were above 0.98 (CI: 0.97-0.99). Although there was a small tendency to lower Ki67 PIs when using VDS (mean difference 0.4%), Bland-Altman plots did not reveal any skewness in any data ranges (see Figure 5.4). Similar results were obtained when comparing the human observer with VDS obtained on the whole core (and not just sampled areas).
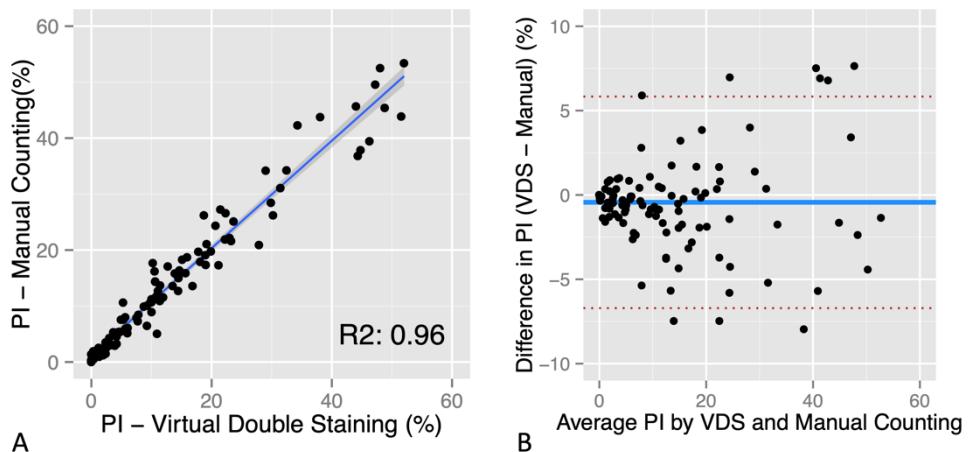
*Figure 5.4. Comparison of Manual Ki67 estimation and Virtual Double Staining in the same sampled areas obtained by Systematic Uniform Random Sampling. **A**. Correlation plot. **B**. Bland-Altman plot. (adapted from Røge R. et al. Proliferation Assessment in Breast Carcinomas Using Digital Image Analysis Based on Virtual Ki67/cytokeratin Double Staining. Breast Cancer Res Treat, 158 (1), 11-19. (study 2))*

Area overlap between neighboring slides where examined in five serial sections stained for PCK. The mean area overlap between

neighboring sections was 91% (range 82%-98%). Overlap decreased by distance between slides.
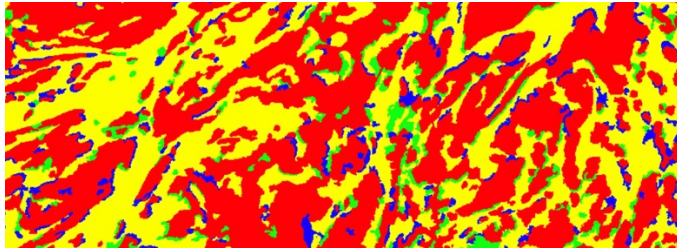
Figure 5.5. Overlap between to neighboring slides stained for PCK. Colors indicates PCK area status in the two slides: Red, PCK-positive in both slides. Yellow, PCK-negative in both slides. Blue/green, PCK-positive in one of the slides. (Adapted from Røge R. et al. Proliferation Assessment in Breast Carcinomas Using Digital Image Analysis Based on Virtual Ki67/cytokeratin Double Staining. Breast Cancer Res Treat, 158 (1), 11-19. (study 2))

When assuming a Ki67 cutoff point of 20%, the human observer classified 74% of the carcinomas as Ki67 Low. In comparison, the VDS-algorithm categorized 75% of the tumours as Ki67 Low.

5.3 Study 3

38 of the 41 included tissue cores were adequate for further analysis (three cores were excluded due to damaged/missing tissue or background staining PCK staining). It was observed that significantly fewer cells were detected when slides were stained on the Ventana Benchmark Ultra platform (mean 4.4% fewer than the overall average).

Differences in Ki67 PIs (percentage points=pp) obtained by comparing the results from the VDS algorithm to the overall average (for each individual core) are presented in Figure 5.6 for the different clone, format and stainer platform combinations.
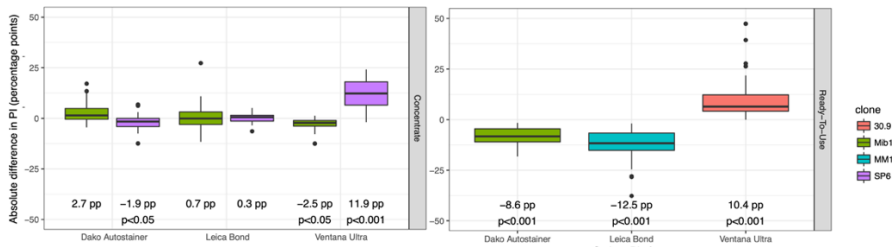
Figure 5.6. Absolute difference (percentage points = pp) in proliferation indices between the most commonly used antibody clones, formats and stainer platforms. The numbers represent the absolute deviation from the overall average. (adapted from Røge et al. Impact of Primary Antibody, Format, and Stainer Platform on Ki67 Proliferation Indices in Breast Carcinomas. Appl Immunohistochem Mol Morphol, 27 (10), 732-739. (study 3)).

The commonly used mAb clone Mib1 as a concentrated format obtained similar Ki67 PIs as the overall average on the Dako Autostainer and Leica Bond platforms. On the Ventana Benchmark Ultra, concentrated Mib1 provided an average Ki67 PI 2.2 pp lower than overall average. The RTU format of Mib1 applied on the Dako Autostainer had 8.6 pp lower Ki67 than the overall average.
Similar results were seen for the concentrated SP6, except on the Ventana Benchmark platform where 11.9 pp higher Ki67 PIs were seen. The MM1 clone, only available as RTU for the Leica Bond stainer, provided the lowest Ki67 PIs – on average 12.5 pp lower than the overall average. The rmAb 30.9, only available as RTU for the Ventana platform, obtained the highest Ki67 PIs namely 10.4 pp higher than the overall average.
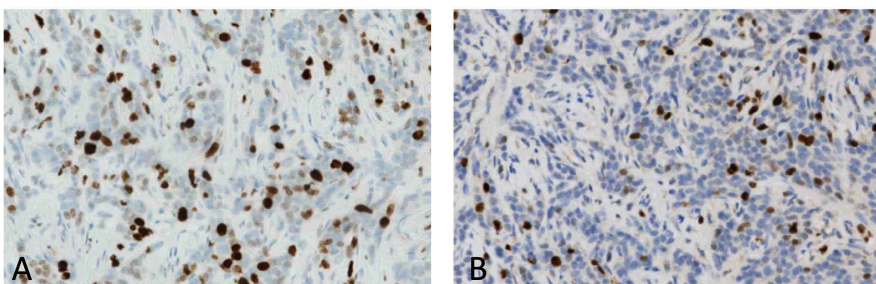
*Figure 5.7. Ki67 stained breast carcinoma using: **A**. Concentrated mAb SP6 on Ventana Benchmark Ultra (overall average Ki67 PI 38%). **B**. RTU MM1 on Leica Bond (overall average Ki67 PI 12%).*

In general, the results showed significant variations in Ki67 PIs between the different clones, formats and stainer platforms.

The study also compared H-scores, which showed similar results as the Ki67 PIs. Using the suggested cutoff of 20% between Ki67 Low or High tumours (Luminal type A and B), the proportion of Ki67 High carcinomas was 63%. If the MM1 RTU products was applied on the Leica Bond platform, 42% were classified as Ki67 High. On the other hand, using either clones 30.9 or SP6 on the Ventana Benchmark platform, 76% of carcinomas were Ki67 High.

5.4 Study 4

The day-to-day variation in H-scores for the three cell lines were low. The coeffiecient of variation was below 5% for all cores and methods (except one core using QuPath) (see Figure 5.8).
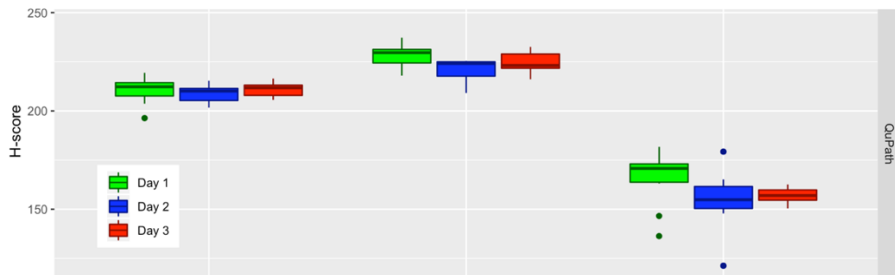
Figure 5.8. Day-to-day variation in H-scores in three cores of cell lines (n=10). H-scores obtained using QuPath. (Adapted from Røge et al. Image Analyses Assessed Cell Lines as Potential Performance Controls of Ki67 Immunstained Slides. Appl Immunhistochem Mol Morphol (Study 4)).

H-scores decreased with increasing dilution of the primary Ab (imitating suboptimal staining protocols). On average, decreasing the Ab concentration to 50%, 25% and 12.5% decreased the H-scores 20.1%, 35.5% and 57.5%, respectively. It was possible to visually select cutoff for each cell lines that separated optimally and suboptimally stained slides (see Figure 5.9).
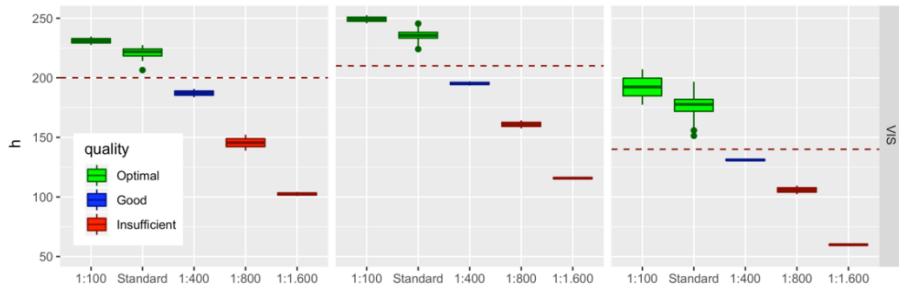
Figure 5.9. H-scores in three cell cultures staining using a range of primary antibody (Mib1) dilutions and quantified using VIS (Visiopharm A/S). Boxplots are coloured depending on the assessed quality of the clinical tissue on the slide. The dashed lines show proposed H-score cutoffs between optimally and suboptimally stained slides. (Adapted from Røge et al. Image Analyses Assessed Cell Lines as Potential Performance Controls of Ki67 Immunstained Slides. Appl Immunhistochem Mol Morphol (Study 4)).

The VIS platform detected 14-18% less nuclei than the QuPath algorithm. There was a tendency to higher H-scores using the VIS DIA program, although this difference was not statistically significant.

6 Discussion

Estimation of Ki67 PIs has still not been widely adopted in clinical diagnostics of breast cancer despite a plethora of studies indicating that Ki67 possesses prognostic and predictive information. The main challenge for implementation of this potential biomarker is the lack of standardization in both Ki67 IHC procedures and interpretation of the results. The general purposes of this PhD project were to elaborate on some of these standardization challenges and examine possible solutions.

We examined the inter-observer variability and methodology for Ki67 PI estimation in a large cohort of human observers (Study 1) and found significant variations in both Ki67 PI results and scoring methodology. We then validated a new DIA algorithm, Virtual Double Staining, for Ki67 PI estimation by comparison to a human observer that used the current stereologically gold standard (Study 2). Additionally, we applied DIA to examine the influence of IHC primary Ab, format and stainer platform on Ki67 PI and found significant differences between the different combinations (study 3). Finally, we examined cell lines assessed by DIA as potential IHC performance controls (Study 4).

6.1 Ki67 estimation in a human cohort (Study 1)

The principal findings of study 1 was significant differences in Ki67 PI obtained by the participants in the same 15 breast cancer cores available as digital images in a webmodule. The participants used different methods for obtaining Ki67 PIs both in terms on how (eyeballing or manual counting of a number of cells) and which area (overall estimate or hotspots) they scored. Participants that counted manually and/or scored in hotspots tended to have higher Ki67 PIs.

There was a good correlation between the Ki67 PIs among the participants (ICC score 0.85), but when applying a cutoff (20%) only moderate kappa-values were achieved. This is in line with other inter-observer studies of Ki67: Shui et al. examined the inter-observer variations among five pathologists estimating Ki67 PIs in 160 breast carcinomas and found an ICC of 0.904 when assessing the overall Ki67 PI average [88]. However, when categorizing in three Ki67 intervals, agreement was significantly lower for tumours close to 20%. Similar results has been reported by other groups [63,89,90]. A recent study by the Danish Breast Cancer Group

(DBCG), found, as we did, higher Ki67 PIs when scoring in hotspots rather than the overall average [91]. As in our study, Focke et al. found significantly higher Ki67 PI when examining in hotspots and counting fewer cells [92].

To the best of our knowledge, this is the largest study that compared the observer's job title (as an indirect measure for experience). We found no significantly different Ki67 PIs and similar variations which somewhat controversially indicated that pathology experience was less important. However, this could also be related to recruited participants that were all contact person in the NordiQC breast cancer module and hence possibly had experience with breast cancer and Ki67 PI estimation regardless of their position. Additionally, the circulated material consisted mostly of ductal carcinomas in which neoplastic cells tended to be easier to identify morphologically than in e.g. lobular carcinomas.

The study used centrally stained slides in order to limit the variability of the Ki67 PI estimation to the interpretation of the IHC stained slide and not variability in IHC assays. However, it was performed digitally which is not the standard microscopy method in most laboratories and hence Ki67 PI could have been influenced by the quality and brightness of local computer monitors. A few studies comparing computer monitors and conventional microscopy found comparable results in mitotic counts and indicate that this may be of minor importance [93,94].

The participants were asked to score the virtual slides using their standard method and had not received any previous guidelines. Our results probably reflect variations in both scoring and scoring methodology in clinical pathology laboratories. Large efforts have been made internationally to develop scoring systems that could standardize the post-analytical phase of the Ki67 IHC stain. The International Ki67 Working Group developed a standardized scoring procedure in TMAs, where 250 cells from both the top and bottom of each core was scored [95]. Sixteen laboratories had been trained in the method before the study. The study achieved an impressive ICC of 0.94. Similar results were later obtained for core-cut biopsies and tissue sections [96,97]. These results indicate that standardized scoring systems may decrease inter-observer variability.

6.2 Digital Image Analysis (Study 1-3)

In study 2, we investigated the applicability of DIA, specifically VDS for Ki67, by comparing results obtained from the algorithm with a human observer estimating Ki67 PI using stereological counting frames in Systemic Uniform Random Sampled areas in 140 breast carcinomas. The ICC for the correlation between the two methods was 0.98 (CI: 0.97-0.99) indicative of an excellent correlation according to Koo et al. [98].

On average the VDS detected 14% fewer cells than the human observer. Qualitative examination of slides revealed that the algorithm sometimes missed small, weakly stained nuclei (DAB and hematoxylin) and nuclei outside the regions of interest that had been excluded due to small misalignments of the slides. Our experiment with agreement of alignment showed that overlap between slides decreased significantly in non-neighbouring slides which affected especially diffusely infiltrating tumours more than solid tumours. It is essential that slides used for VDS are neighboring slides which may require an update laboratory workflow. Even in well aligned slides one must expect small changes in the neoplastic area which will include stroma cells in the regions of interest and hence dilute Ki67 PI with stromal cells (that most commonly have lower Ki67 PI).

Normal glandular epithelium and carcinoma in situ (CIS) are also positive for PCK and would be included in the tumour regions of interest. In this study, we selected tumour areas without normal epithelium and CIS, but this is often not possible in clinical settings, which could prevent unsupervised analysis. One possible solution for this could be inclusion of basal cell markers (e.g. p63 and/or heavy chain smooth muscle myosin) in a virtual triple stain, where glands with positive staining for basal markers are excluded.

In the same manner, tumour infiltrating lymphocytes (TILs) will be included in the regions of interest if located among the neoplastic epithelial cells. If the number of lymphocytes is high and they have a Ki67 PI deviating from the neoplastic cells they could influence they final result. Especially medullary carcinomas have high Ki67 PIs and contain many lymphocytes [99]. We had included two medullary carcinomas in this study. The human observer scored higher Ki67 PI in one of the cases and lower in the other. Hence, the question on influence of TIL remains and warrants further studies for clarification.

The literature shows that many well-functioning algorithms for estimation of Ki67 PIs are available. The main drawback with many of these algorithms is lack of tumour recognition and requirement for manual selection of regions of interest. This is time consuming especially for whole slide sections. One group addressed this by using physical double stains in melanomas (Mart1/Ki67) and breast cancer (KL1/Ki67) [100,101]. However, the authors found lower Ki67 PI in physcial double stains when comparing to single stains which may be due to overlapping of the chromogens.

One limitation to study 2 is the comparison with a single human observer. Since the human observer (RR) participated in the development of the algorithm, this may be optimized to his interpretation of Ki67 positive and negative cells. For this reason, it would be important to validate the algorithm to a larger cohort of observers.

To achieve this, we applied the VDS algorithm on the digital slides scored by the participants (that scored an overall average) in Study 1 and found that VDS Ki67 PIs were similar to the mean (see Figure 5.3). Although this is a crude comparison, wisdoms of crowds has been a well-established concept since Francis Galton's famous experiment at a livestock fair, where the median of the attendees' guesses correctly predicted the weight of an ox [102].

The VDS approach has since publication of Study 2 also been examined in whole slide images of breast carcinoma using two different platforms (Visiopharm and Indica labs). The authors found good correlations with manual counting on both platforms [103]. As in our study, a rather large number of tissues had to be excluded due to folding of tissue which puts greater demands on laboratories for optimal slides and require special training of laboratory technicians. Additionally, a large study examining accuracy in detection of PAM50 validated Luminal B subtypes showed that VDS outperformed human observers [104]. The authors examined VDS estimated Ki67 PI as an overall average both also in hot-spots and in the invasive margin and found that hot spots had slightly better prognostic value. In a recent study, the group confirmed that Ki67 PI estimated by VDS outperformed manual observers (especially in hot spots) in classification of luminal type and with regard to prediction of overall and relapse-free survival [65].

Study 2 did not examine Ki67 in hot spots or the invasive margin but rather the overall average. There is no reason to expect that our

validation of the segmentation of Ki67-positive and negative nuclei cannot be extrapolated to estimation in invasive margins or hot spots. Hot spots are detected mathematically after the nuclear segmentation [105] and invasive margins are based on a manual selection of ROI.

In study 3 we examined the impact of different Ki67 mAb clones, clone formats and stainer platforms on Ki67 PI and H-scores in 41 breast carcinomas obtained by VDS. The results for Ki67 PIs and H-scores were relatively similar. The principal finding was significantly different Ki67 PI when using different primary Ab clones or formats. Additionally, results were platform related for some clones.

The results indicated that Ki67s are comparable when using the concentrated format of Mib1 and SP6 on the Dako Autostainer or Leica Bond (using optimized protocols). On the Ventana Benchmark platform, the Mib1 clone provided lower PIs, while the mAb clones SP6 and 30.9 gave higher results. This could potentially be related to the sensitivity of the individual visualization system to different Ab types. SP6 and 30.9 are both raised in rabbits, while Mib1 is a mouse monoclonal Ab. The RTU format of the Mib1 provided lower Ki67 PIs, indicating that the overall sensitivity of the RTU system and clone is lower than in-house optimized protocols. This could perhaps be related to the Ab concentration in the RTU product.

Our results are in line with other studies comparing Ki67 IHC Ab clones: Fasanella et al. compared the SP6 and MM1 antibodies stained on Dako Autostainer 720 and Leica Bond Max [106]. The authors found significantly higher Ki67-scores when using SP6 although the antibodies were not directly compared. Instead, two consecutive breast carcinomas cohorts were stained using SP6 or MM1 and summary statistics compared. Another group compared the Mib1 and MM1 clones in normal and neoplastic tissues and found, as in study 3, that the Mib1 clone provided significantly higher Ki67 PIs [107,108]. Both studies are rather old and the IHC was performed on a now outdated machine with obsolete visualization systems, which makes comparison to our study less relevant. Another study by Ekholm et al. examined the prognostic value of Mib1 and SP6 Ab clone and found that clones were comparable in detecting distant disease-free survival [109]. As in the previous studies, the IHC was performed on an older platform with an outdated biotin-avidin detection system. One group also compared the Mib1 and SP6 clones and their ability to predict progression of advanced disease and found comparable results for

the clones [110]. The authors also concluded that the Mib1 clone was less suited for DIA due to significant background staining. We did not experience such background issue and suspect that the background staining in more likely related suboptimal staining protocols rather than the Mib1 clone itself.

Focke et al. compared thirty European laboratories that stained serial sections from the same breast carcinoma TMA using their in-house Ki67 protocol [90]. Slides were assessed centrally by one human observer. Significant differences were observed: The proportion of luminal type A varied from 17-51%. The variance remained when splitting up the laboratories depending on the primary Ab clone applied. This is most likely caused by suboptimal protocols being employed by some of laboratories. Although the study incorporated "participation in an external QA program" as a parameter, a better parameter for future studies could perhaps be "performance in external QA programs". Only optimized protocols were applied in our study.

All aspects of Ki67 estimation from preanalytics to IHC staining and interpretation must be considered when trying to optimize the Ki67 test. Going through the literature, enormous efforts have been made to examine and expand on the interpretation. Conversely, only relative few studies have been made concerning the Ki67 IHC protocols and preanalytics. In our opinion, significantly greater efforts should be devoted to developing standardized IHC methods that no doubt will ease interpretation.

We observed a significantly lower number of cells detected on the Ventana Benchmark platform. This could be related to either smaller areas detected or fewer nuclei. Data showed that the PCK-stained areas were similar in size across the different platform. If the algorithm systematically detected fewer negative cells (due to weaker counterstaining), this could potentially explain some increases in the Ki67 PIs on the Ventana platform. However, qualitative assessment of the Ki67 stained slides did not reveal any platform related drops in cell counts. Theoretical calculations showed that a loss of 10% negative cell on average only would increase the Ki67 PI by 0-3%. We do not find that the increased PIs in the two clones 30.9 and SP6 on the Ventana platform can be explained by this phenomenon. Especially when the Mib1 clone yielded lower PIs.

Although we could not provide any explanation for the difference in cell counts, it is most likely related to small difference in the staining

intensity of both DAB and hematoxylin. This underlines the necessity of calibrations of DIA algorithms. In this study, it was not possible to optimize the algorithm for each stainer platform and visualization system that uses hematoxylin and DAB with nuance differences in staining color. Internal data from a DIA experiment of slides stained by the NordiQC participants showed that the wide variety of intensities in DAB and hematoxylin prevented analysis. In recognition of this, it has since become possible to adjust the VDS algorithm to laboratory specific intensities by use of reference materials.

The color calibration of the image capture devices as slide scanner can also impart the outcome of DIA algorithms. In the context of this project, slides (a TMA with 41 breast carcinomas) were scanned using different color profiles on the scanner before and after color calibration of the scanner (data not presented). Although no statistical significance could be obtained, there was a tendency towards different results before and after calibration. The potential source of error would not affect the results in this thesis, since all slides for an experiment were scanned in one batch. However, it is necessary to take this in account when comparing digital slides from different institutions, scanner manufacturers and achieved material. This problem has also been recorded in the literature and is potentially solvable using on slide color controls [111].

6.3 Immunohistochemistry controls (Study 1 and 4)

In study 1, the most common (48%) control tissue among the NordiQC participants was lymphoid tissue (tonsil, lymph node or appendix), while a rather large group (38%) stated that they did not use control tissue or internal controls. This finding is not surprising since it well known that many clinical pathology laboratories still do not use on-slide controls. IHC reactions without on-slide controls, will, in the potential absence of internal positive cells, prevent the pathologist from knowing if the IHC reaction succeeded.

In study 4, we showed that cell lines stained for Ki67 and assessed by DIA provided rather stable day-to-day H-scores. Additionally, H-scores reflected the technical sensitivity in assays with simulated suboptimal protocol parameters. This allowed identification of

cutoffs in H-scores that reliably identified suboptimal staining protocols.

To the best of our knowledge, this is the first study that examined the potential of cell lines as performance controls in Ki67 immunostained slides. Cell lines engineered to express certain epitopes are often used in academical laboratories as controls. However, this method is more seldomly used in laboratories with access to clinical tissues. Previously, cell lines have been suggested as performance controls for HER2, although this method has not yet been incorporated in daily practices [112]. The authors identified cell lines with (and without) HER2 gene amplification and corresponding 3+ and 2+ IHC reaction.

When selecting controls for IHC experiment it is imperative that tissues include the clinical relevant range of epitope expression [24,113]. Tissue with very high or very low levels of epitope expression will provide only little information about the sensitivity and successfulness of the reaction. The cell lines examined in study 4 consisted of cells with varying Ki67 expression ranging from weak to strong.
Cell lines cannot be used alone as external IHC controls, since they would only monitor the sensitivity, and must be supplemented with other tissue for specificity and preanalytics.

One extreme data point obtained using QuPath (core C) was below the cutoff. Visual inspection of this core revealed only strongly positive and negatively stained cells indicating that the IHC reaction to some degree had failed. Score obtained on the second platform for this core was also low (but not below the cut-off value).

One of the strengths with this study was the use of two different DIA platforms (one commercial and one open source) for obtaining H-scores. Although one platform systematically identified more cells than the other (and reported lower H-scores), there was good correlations between the results obtained by the two platforms. If the cell lines and DIA were applied in a clinical laboratory as a clinical sensitivity surveillance system, deviations from predetermined H-scores range would be the primary benchmark. Using both platforms, it was possible to identify H-score cut-offs that could identify suboptimally stained slides. For this reason, both platforms can be used. Previously, we examined the potential of using cell lines as quality indicators in a NordiQC Ki67 run using a custom developed algorithm [114]. The algorithm showed great

promise in a test set (forty cases), but this could not be extrapolated to the whole cohort due to significant differences in color intensities. This indicates that the cell lines and DIA H-score as sensitivity surveillance system may be limited to intra-laboratory surveillance and not inter-laboratory comparisons.

One limit of the study was the rather small number of samples examined. This proof of concept should be validated by further studies. The method could potentially be expanded to encompass other nuclear markers as ER, PR, Sox10, Bcl-6 and Pax8 provided that relevant cell lines with clinically relevant expression ranges can be found.

7 Conclusions

In this PhD study we examined the current practices on Ki67 scoring in breast cancer in a large cohort of clinical pathology laboratories and found significant differences. Additionally, we saw variations in Ki67 scoring depending on the methods used. This underlines the need for standardization since variations may impact the treatment decision in patients.

DIA for obtaining Ki67 PIs is possibly one such standardization step. We validated the VDS algorithm to a human observer and found good correlation between Ki67 PIs obtained by the human observer and DIA. After further validations, this algorithm could possible by applied in clinical practice.

We applied the algorithm to compare Ki67 PIs in clinical tissues stained with different antibodies, formats and stainer platforms and found significant differences in Ki67 PIs when applying different combinations. One major focus of Ki67 in breast cancer has been the interpretation of stained slides. Our results show that more attention should be given to the IHC analysis itself, since the variations introduced in this step is just as large as by interpretation.

Finally, we introduce a novel method for monitoring the performance sensitivity of Ki67 IHC stained slides. Using cell lines and DIA it was possible to detect suboptimally stained slides.

8 Future perspectives

The studies included in this thesis have generated questions for further research.

The large variation in scoring methods observed in study calls for international standardization of counting methods. Likewise, the question on where to score remains. Some studies indicate that Ki67 hot spots provide better predictive information than the overall average. However, definitions on hot spots vary and should be harmonized in coming studies validating this.

The VDS algorithm has also been examined by other groups that proved that the algorithm worked on whole slide tissue and was better in predictions of survival [65,103,104]. However, it would be interesting to see comparisons of VDS applied by different laboratories on both the same IHC slides and slides stained in house. Although manual counting of Ki67 PIs is a laborious and time-consuming task, feasibility studies examining the time and cost for implementing and performing VDS are needed. At the moment it would require additional laboratory steps but with the advent of digital pathology platforms this complexity might be reduced.

Especially relevant for DIA, there still do not exist a clear definition of an (Ki67) IHC-positive nuclei. This is imperative for development of DIA systems and may require international consensus definitions developed by pathologist and technologists. Possibly comparative studies could examine prognostic and predictive values of DIA with different definitions of positivity.

At the moment, artificial intelligence is gaining momentum in many scientific fields including DIA. It has been shown many times that well-trained machine learning algorithms can identify tumour tissue in hematoxylin stained slides. This new technique could potentially make VDS obsolete, since tumour identification can be obtained on the Ki67 stained slide. However, until this that been validated VDS could also serve as an unsupervised tool for identification of tumour tissue for training of machine learning algorithms.

The large impact of Ki67 primary Ab clone, format and stainer platform indicates that the source of variations should be explored further. Large studies comparing the effects of these parameters on predictive value of Ki67 and comparison to gene expression assays

should be completed as soon as possible, since the lack of standardization in this field hinders comparison of results from studies using different methodology.

Our cell line approach to monitor IHC performance should be validated in a larger study. One possible study could be to include cell lines on all Ki67 slides in diagnostic laboratory alongside several cores of breast carcinoma and assess the H-scores using DIA.

9 Literature

[1]   Schonk DM, Kuijpers HJ, van Drunen E, van Dalen CH, Geurts van Kessel AH, Verheijen R, et al. Assignment of the gene(s) involved in the expression of the proliferation-related Ki-67 antigen to human chromosome 10. Hum Genet 1989;83:297–9. doi:10.1007/bf00285178.

[2]   Gerdes J, Schwab U, Lemke H, Stein H. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. Int J Cancer 1983;31:13–20. doi:10.1002/ijc.2910310104.

[3]   Gerdes J, Li L, Schlueter C, Duchrow M, Wohlenberg C, Gerlach C, et al. Immunobiochemical and molecular biologic characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. Am J Pathol 1991;138:867–73.

[4]   Booth DG, Earnshaw WC. Ki-67 and the Chromosome Periphery Compartment in Mitosis. Trends Cell Biol 2017;27:906–16. doi:10.1016/j.tcb.2017.08.001.

[5]   Schlüter C, Duchrow M, Wohlenberg C, Becker MH, Key G, Flad HD, et al. The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. J Cell Biol 1993;123:513–22. doi:10.1083/jcb.123.3.513.

[6]   Hofmann K, Bucher P. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. Trends Biochem Sci 1995;20:347–9. doi:10.1016/s0968-0004(00)89072-6.

[7]   Durocher D, Jackson SP. The FHA domain. FEBS Lett 2002;513:58–66. doi:10.1016/S0014-5793(01)03294-X.

[8]   Booth DG, Takagi M, Sanchez-Pulido L, Petfalski E, Vargiu G, Samejima K, et al. Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. Elife 2014;2014:1–22. doi:10.7554/eLife.01641.001.

[9]   Kim H-S, Fernandes G, Lee C-W. Protein Phosphatases Involved in Regulating Mitosis: Facts and Hypotheses. Mol Cells 2016;39:654–62. doi:10.14348/molcells.2016.0214.

[10]   MacCallum DE, Hall PA. Biochemical Characterization of pKi67 with the Identification of a Mitotic-Specific Form Associated with Hyperphosphorylation and Altered DNA Binding. Exp Cell Res 1999;252:186–98. doi:10.1006/excr.1999.4600.

[11] Takagi M, Matsuoka Y, Kurihara T, Yoneda Y. Chmadrin: A novel Ki-67 antigen-related perichromosomal protein possibly implicated in higher order chromatin structure. J Cell Sci 1999;112:2463–72.

[12] Sobecki M, Mrouj K, Camasses A, Parisis N, Nicolas E, Llères D, et al. The cell proliferation antigen Ki-67 organises heterochromatin. Elife 2016;5:1–33. doi:10.7554/eLife.13722.

[13] Lomberk G, Wallrath L, Urrutia R. The Heterochromatin Protein 1 family. Genome Biol 2006;7:228. doi:10.1186/gb-2006-7-7-228.

[14] Cuylen S, Blaukopf C, Politi AZ, Müller-Reichert T, Neumann B, Poser I, et al. Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. Nature 2016;535:308–12. doi:10.1038/nature18610.

[15] Booth DG, Earnshaw WC. Ki-67 and the Chromosome Periphery Compartment in Mitosis. Trends Cell Biol 2017;27:906–16. doi:10.1016/j.tcb.2017.08.001.

[16] Sun X, Bizhanova A, Matheson TD, Yu J, Zhu LJ, Kaufman PD. Ki-67 Contributes to Normal Cell Cycle Progression and Inactive X Heterochromatin in p21 Checkpoint-Proficient Human Cells. Mol Cell Biol 2017;37:1–27. doi:10.1128/MCB.00569-16.

[17] Sobecki M, Mrouj K, Colinge J, Gerbe F, Jay P, Krasinska L, et al. Cell-cycle regulation accounts for variability in Ki-67 expression levels. Cancer Res 2017;77:2722–34. doi:10.1158/0008-5472.CAN-16-0707.

[18] Kill IR. Localisation of the Ki-67 antigen within the nucleolus. Evidence for a fibrillarin-deficient region of the dense fibrillar component. J Cell Sci 1996;109 (Pt 6:1253–63.

[19] Hall PA, Richards MA, Gregory WM, D'Ardenne AJ, Lister TA, Stansfeld AG. The prognostic value of Ki67 immunostaining in non-Hodgkin's lymphoma. J Pathol 1988;154:223–35. doi:10.1002/path.1711540305.

[20] Veronese SM, Gambacorta M, Gottardi O, Scanzi F, Ferrari M, Lampertico P. Proliferation index as a prognostic marker in breast cancer. Cancer 1993;71:3926–31. doi:10.1002/1097-0142(19930615)71:12<3926::aid-cncr2820711221>3.0.co;2-2.

[21] Cattoretti G, Becker MH, Key G, Duchrow M, Schlüter C, Galle J, et al. Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave-processed formalin-fixed

paraffin sections. J Pathol 1992;168:357–63. doi:10.1002/path.1711680404.

[22] Cattoretti G, Pileri S, Parravicini C, Becker MH, Poggi S, Bifulco C, et al. Antigen unmasking on formalin-fixed, paraffin-embedded tissue sections. J Pathol 1993;171:83–98. doi:10.1002/path.1711710205.

[23] NordiQC. NordiQC Ki67 run B22, Breast cancer module 2016:https://www.nordiqc.org/downloads/assessments/84_1.

[24] Torlakovic EE, Nielsen S, Francis G, Garratt J, Gilks B, Goldsmith JD, et al. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. Appl Immunohistochem Mol Morphol AIMM 2015;23:1–18. doi:10.1097/PAI.0000000000000163.

[25] Bernard WS, Christopher PW. World cancer report 2014. World Heal Organ 2014:630.

[26] Danckert B, Ferlay J, Engholm G , Hansen HL, Johannesen TB, Khan S, Køtlum JE, Ólafsdóttir E, Schmidt LKH VA and SH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.2 (26.03.2019). 2020:http://www.ancr.nu, accessed 02-jan-2020.

[27] Sinn HP, Kreipe H. A brief overview of the WHO classification of breast tumors, 4th edition, focusing on issues and updates from the 3rd edition. Breast Care 2013;8:149–54. doi:10.1159/000350774.

[28] Reed MEMC, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: Morphology, biomarkers and 'omics. Breast Cancer Res 2015;17:1–11. doi:10.1186/s13058-015-0519-x.

[29] BLOOM HJ, RICHARDSON WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359–77. doi:10.1038/bjc.1957.43.

[30] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403–10. doi:10.1111/j.1365-2559.1991.tb00229.x.

[31] Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast carcinoma. J Clin Oncol 2008;26:3153–8. doi:10.1200/JCO.2007.15.5986.

[32] DBCG. Kirurgisk behandling af brystkræft 2019.

http://www.dmcg.dk/siteassets/forside/kliniske-retningslinjer/godkendte-kr/dbcg_kirurgisk-behandling_adm-godk_jbh160919.pdf.

[33] DBCG. Medicinsk behandling af brystkræft 2019.

[34] Alexieva-Figusch J, Van Putten WL, Blankenstein MA, Blonk-Van Der Wijst J, Klijn JG. The prognostic value and relationships of patient characteristics, estrogen and progestin receptors, and site of relapse in primary breast cancer. Cancer 1988;61:758–68. doi:10.1002/1097-0142(19880215)61:4<758::aid-cncr2820610421>3.0.co;2-t.

[35] Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 1998;11:155–68.

[36] Baum M, Brinkley DM, Dossett JA, McPherson K, Patterson JS, Rubens RD, et al. Improved survival among patients treated with adjuvant tamoxifen after mastectomy for early breast cancer. Lancet (London, England) 1983;2:450. doi:10.1016/s0140-6736(83)90406-3.

[37] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet (London, England) n.d.;365:1687–717. doi:10.1016/S0140-6736(05)66544-0.

[38] Kos Z, Dabbs DJ. Biomarker assessment and molecular testing for prognostication in breast cancer. Histopathology 2016;68:70–85. doi:10.1111/his.12795.

[39] Bardou V-J, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol 2003;21:1973–9. doi:10.1200/JCO.2003.09.099.

[40] Dowsett M, Houghton J, Iden C, Salter J, Farndon J, A'Hern R, et al. Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status. Ann Oncol Off J Eur Soc Med Oncol 2006;17:818–26. doi:10.1093/annonc/mdl016.

[41] Hefti MM, Hu R, Knoblauch NW, Collins LC, Haibe-Kains B, Tamimi RM, et al. Estrogen receptor negative/progesterone receptor positive breast cancer is not a reproducible subtype. Breast Cancer Res 2013;15:1. doi:10.1186/bcr3462.

[42] Köninki K, Tanner M, Auvinen A, Isola J. HER-2 positive

breast cancer: Decreasing proportion but stable incidence in Finnish population from 1982 to 2005. Breast Cancer Res 2009;11:1–6. doi:10.1186/bcr2322.

[43]   Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987;235:177–82. doi:10.1126/science.3798106.

[44]   Cooke T, Reeves J, Lanigan A, Stanton P. HER2 as a prognostic and predictive marker for breast cancer. Ann Oncol 2001;12:23–8. doi:10.1093/annonc/12.suppl_1.S23.

[45]   Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 2011;365:1273–83. doi:10.1056/NEJMoa0910383.

[46]   de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer 2007;96:1504–13. doi:10.1038/sj.bjc.6603756.

[47]   Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients. Breast 2008;17:323–34. doi:10.1016/j.breast.2008.02.002.

[48]   Wu Q, Ma G, Deng Y, Luo W, Zhao Y, Li W, et al. Prognostic Value of Ki-67 in Patients With Resected Triple-Negative Breast Cancer: A Meta-Analysis. Front Oncol 2019;9:1068. doi:10.3389/fonc.2019.01068.

[49]   Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52. doi:10.1038/35021093.

[50]   Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74. doi:10.1073/pnas.191367098.

[51]   Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23. doi:10.1073/pnas.0932692100.

[52]   Curtis, C. et al. The Cancer Genome Atlas Network. Nature 2012;490:61–70. doi:10.1038/nature11412.Comprehensive.

[53]   Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J,

et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 2009;101:736–50. doi:10.1093/jnci/djp082.

[54] Nielsen TO, Jensen MB, Burugu S, Gao D, Tykjaer Jørgensen CL, Balslev E, et al. High-risk premenopausal Luminal A breast cancer patients derive no benefit from adjuvant cyclophosphamide-based chemotherapy: Results from the DBCG77B clinical trial. Clin Cancer Res 2017;23:946–53. doi:10.1158/1078-0432.CCR-16-1278.

[55] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26. doi:10.1056/NEJMoa041588.

[56] Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7. doi:10.1200/JCO.2008.18.1370.

[57] Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 2015;26:1533–46. doi:10.1093/annonc/mdv221.

[58] Regan MM, Pagani O, Walley B, Torrisi R, Perez EA, Francis P, et al. Premenopausal endocrine-responsive early breast cancer: who receives chemotherapy? Ann Oncol Off J Eur Soc Med Oncol 2008;19:1231–41. doi:10.1093/annonc/mdn037.

[59] Goldhirsch a, Wood WC, Coates a S, Gelber RD, Thürlimann B, Senn H-J. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 2011;22:1736–47. doi:10.1093/annonc/mdr304.

[60] Goldhirsch a, Winer EP, Coates a S, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013;24:2206–23. doi:10.1093/annonc/mdt303.

[61] Curigliano G, Burstein HJ, P Winer E, Gnant M, Dubsky P, Loibl S, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of

Early Breast Cancer 2017. Ann Oncol Off J Eur Soc Med Oncol 2017;28:1700–12. doi:10.1093/annonc/mdx308.

[62] Burstein HJ, Curigliano G, Loibl S, Dubsky P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Ann Oncol Off J Eur Soc Med Oncol 2019;30:1541–57. doi:10.1093/annonc/mdz235.

[63] Polley M-YC, Leung SCY, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013;105:1897–906. doi:10.1093/jnci/djt306.

[64] Jang MH, Kim HJ, Chung YR, Lee Y, Park SY. A comparison of Ki-67 counting methods in luminal Breast Cancer: The Average Method vs. the Hot Spot Method. PLoS One 2017;12:1–15. doi:10.1371/journal.pone.0172031.

[65] Stålhammar G, Robertson S, Wedlund L, Lippert M, Rantalainen M, Bergh J, et al. Digital image analysis of Ki67 in hot spots is superior to both manual Ki67 and mitotic counts in breast cancer. Histopathology 2018;72:974–89. doi:10.1111/his.13452.

[66] Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. J Pathol Inform 2019;10:9. doi:10.4103/jpi.jpi_82_18.

[67] Grunkin M, Raundahl J, Foged NT. Practical considerations of image analysis and quantification of signal transduction IHC staining. Methods Mol Biol 2011;717:143–54. doi:10.1007/978-1-61779-024-9_8.

[68] Sucaet Y, Waelput W. Digital Pathology. Cham: Springer International Publishing; 2014. doi:10.1007/978-3-319-08780-1.

[69] Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. Anal Quant Cytol Histol 1992;14:459–73.

[70] Serag A, Ion-Margineanu A, Qureshi H, McMillan R, Saint Martin M-J, Diamond J, et al. Translational AI and Deep Learning in Diagnostic Pathology. Front Med 2019;6. doi:10.3389/fmed.2019.00185.

[71] Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology Image Analysis Using Segmentation Deep Learning Algorithms. Am J Pathol 2019;189:1686–98. doi:10.1016/j.ajpath.2019.05.007.

[72] Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer 2019;118:91–6. doi:10.1016/j.ejca.2019.06.012.

[73] Tuominen VJ, Ruotoistenmäki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. Breast Cancer Res 2010;12:R56. doi:10.1186/bcr2615.

[74] Yeo MK, Kim HE, Kim SH, Chae BJ, Song BJ, Lee A. Clinical usefulness of the free web-based image analysis application ImmunoRatio for assessment of Ki-67 labelling index in breast cancer. J Clin Pathol 2017;70:715–9. doi:10.1136/jclinpath-2016-204162.

[75] Fulawka L, Halon A. Proliferation index evaluation in breast cancer using imagej and immunoratio applications. Anticancer Res 2016;36:3965–72.

[76] Volynskaya Z, Mete O, Pakbaz S, Al-Ghamdi D, Asa S. Ki67 quantitative interpretation: Insights using image analysis. J Pathol Inform 2019;10:8. doi:10.4103/jpi.jpi_76_18.

[77] Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R. A comparison of visual assessment and automated digital image analysis of Ki67 labeling index in breast cancer. PLoS One 2016;11:1–11. doi:10.1371/journal.pone.0150505.

[78] Tewary S, Arun I, Ahmed R, Chatterjee S, Chakraborty C. SmartIHC-Analyzer: smartphone assisted microscopic image analytics for automated Ki-67 quantification in breast cancer evaluation. Anal Methods 2017;9:6161–70. doi:10.1039/C7AY02302B.

[79] Yaziji H, Taylor CR, Goldstein NS, Dabbs DJ, Hammond EH, Hewlett B, et al. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. Appl Immunohistochem Mol Morphol AIMM 2008;16:513–20. doi:10.1097/PAI.0b013e31818a9d3a.

[80] Torlakovic EE, Cheung CC, D'Arrigo C, Dietel M, Francis GD, Gilks CB, 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 AIMM 2017;25:79–85. doi:10.1097/PAI.0000000000000444.

[81] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep 2017;7:16878. doi:10.1038/s41598-017-17204-5.

[82] Gundersen HJ, Bendtsen TF, Korbo L, Marcussen N, Møller A, Nielsen K, et al. Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. APMIS 1988;96:379–94.

[83] Vyberg M, Nielsen S. Proficiency testing in immunohistochemistry--experiences from Nordic Immunohistochemical Quality Control (NordiQC). Virchows Arch 2016;468:19–29. doi:10.1007/s00428-015-1829-1.

[84] Simons JE, Holmes DT. Reproducible Research and Reports with R 2019:2018–20. doi:10.1373/jalm.2018.028373.

[85] Wilkinson L. The Grammar of Graphics. New York: Springer-Verlag; 2005. doi:10.1007/0-387-28695-0.

[86] McGill R, Tukey JW, Larsen WA. Variations of Box Plots. Am Stat 1978;32:12. doi:10.2307/2683468.

[87] Altman DG, Bland JM. Measurement in Medicine : the Analysis of Method Comparison Studies. Stat 1983;32:307–17. doi:10.2307/2987937.

[88] Shui R, Yu B, Bi R, Yang F, Yang W. An interobserver reproducibility analysis of Ki67 visual assessment in breast cancer. PLoS One 2015;10:e0125131. doi:10.1371/journal.pone.0125131.

[89] Chung YR, Jang MH, Park SY, Gong G, Jung WH, Kwon SY, et al. Interobserver variability of Ki-67 measurement in breast cancer. J Pathol Transl Med 2016;50:129–37. doi:10.4132/jptm.2015.12.24.

[90] Focke CM, Bürger H, van Diest PJ, Finsterbusch K, Gläser D, Korsching E, et al. Interlaboratory variability of Ki67 staining in breast cancer. Eur J Cancer 2017;84:219–27. doi:10.1016/j.ejca.2017.07.041.

[91] Laenkholm A-V, Grabau D, Møller Talman M-L, Balslev E, Bak Jylling AM, Tabor TP, et al. An inter-observer Ki67 reproducibility study applying two different assessment methods: on behalf of the Danish Scientific Committee of Pathology, Danish breast cancer cooperative group (DBCG). Acta Oncol 2018;57:83–9. doi:10.1080/0284186X.2017.1404127.

[92] Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. Breast Cancer Res Treat 2016;159:257–63. doi:10.1007/s10549-016-3950-5.

[93] Norgan AP, Suman VJ, Brown CL, Flotte TJ, Mounajjed T. Comparison of a medical-grade monitor vs commercial off-the-shelf display for mitotic figure enumeration and small object (Helicobacter pylori) detection. Am J Clin Pathol

2018;149:181–5. doi:10.1093/AJCP/AQX154.

[94]   Wei B-R, Halsey CH, Hoover SB, Puri M, Yang HH, Gallas BD, et al. Agreement in Histological Assessment of Mitotic Activity Between Microscopy and Digital Whole Slide Images Informs Conversion for Clinical Diagnosis. Acad Pathol 2019;6:237428951985984. doi:10.1177/2374289519859841.

[95]   Polley M-YC, Leung SCY, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JMS, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015;28:778–86. doi:10.1038/modpathol.2015.38.

[96]   Leung SCY, Nielsen TO, Zabaglo L, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. NPJ Breast Cancer 2016;2:16014. doi:10.1038/npjbcancer.2016.14.

[97]   Leung SCY, Nielsen TO, Zabaglo LA, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. Histopathology 2019;75:225–35. doi:10.1111/his.13880.

[98]   Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155–63. doi:10.1016/j.jcm.2016.02.012.

[99]   Alvarenga CA, Paravidino PI, Alvarenga M, Gomes M, Dufloth R, Zeferino LC, et al. Reappraisal of immunohistochemical profiling of special histological types of breast carcinomas: a study of 121 cases of eight different subtypes. J Clin Pathol 2012;65:1066–71. doi:10.1136/jclinpath-2012-200885.

[100]   Nielsen PS, Riber-Hansen R, Raundahl J, Steiniche T. Automated quantification of MART1-verified Ki67 indices by digital image analysis in melanocytic lesions. Arch Pathol Lab Med 2012;136:627–34. doi:10.5858/arpa.2011-0360-OA.

[101]   Nielsen PS, Bentzer NK, Jensen V, Steiniche T, Jylling AMB. Immunohistochemical Ki-67/KL1 double stains increase accuracy of Ki-67 indices in breast cancer and simplify automated image analysis. Appl Immunohistochem Mol Morphol AIMM 2014;22:568–76. doi:10.1097/PAI.0b013e3182a84b99.

[102]   Galton F. Vox populi. Nature 1907;75:450–1.

[103] Koopman T, Buikema HJ, Hollema H, de Bock GH, van der Vegt B. Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. Breast Cancer Res Treat 2018;169:33–42. doi:10.1007/s10549-018-4669-2.

[104] Stålhammar G, Fuentes Martinez N, Lippert M, Tobin NP, Mølholm I, Kis L, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. Mod Pathol 2016;29:318–29. doi:10.1038/modpathol.2016.34.

[105] Wessel Lindberg A-S, Conradsen K, Larsen R, Friis Lippert M, Røge R, Vyberg M. Quantitative tumor heterogeneity assessment on a nuclear population basis. Cytom Part A 2017;91. doi:10.1002/cyto.a.23047.

[106] Fasanella S, Leonardi E, Cantaloni C, Eccher C, Bazzanella I, Aldovini D, et al. Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies. Diagn Pathol 2011;6 Suppl 1:S7. doi:10.1186/1746-1596-6-S1-S7.

[107] Lindboe CF, Torp SH. Comparison of Ki-67 equivalent antibodies. J Clin Pathol 2002;55:467–71.

[108] Lindboe CF, von der Ohe G, Torp SH. Determination of proliferation index in neoplasms using different Ki-67 equivalent antibodies. Apmis 2003;111:567–70. doi:10.1034/j.1600-0463.2003.1110505.x.

[109] Ekholm M, Beglerbegovic S, Grabau D, Lövgren K, Malmström P, Hartman L, et al. Immunohistochemical assessment of Ki67 with antibodies SP6 and MIB1 in primary breast cancer: a comparison of prognostic value and reproducibility. Histopathology 2014;2:1–9. doi:10.1111/his.12392.

[110] Zabaglo L, Salter J, Anderson H, Quinn E, Hills M, Detre S, et al. Comparative validation of the SP6 antibody to Ki67 in breast cancer. J Clin Pathol 2010;63:800–4. doi:10.1136/jcp.2010.077578.

[111] Yagi Y. Color standardization and optimization in Whole Slide Imaging. Diagn Pathol 2011;6:S15. doi:10.1186/1746-1596-6-S1-S15.

[112] Rhodes A, Jasani B, Couturier J, McKinley MJ, Morgan JM, Dodson AR, et al. A formalin-fixed, paraffin-processed cell line standard for quality control of immunohistochemical assay of HER-2/neu expression in breast cancer. Am J Clin Pathol 2002;117:81–9. doi:10.1309/4NCM-QJ9W-QM0J-6QJE.

[113] Torlakovic EE, Nielsen S, Vyberg M, Taylor CR. Getting controls under control: the time is now for immunohistochemistry. J Clin Pathol 2015;68:879–82. doi:10.1136/jclinpath-2014-202705.

[114] Lanng MB, Møller CB, Andersen ASH, Pálsdóttir ÁA, Røge R, Østergaard LR, et al. Quality assessment of Ki67 staining using cell line proliferation index and stain intensity features. Cytom Part A 2018. doi:10.1002/cyto.a.23683.

10 Appendix: Published papers

Study 1.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Ki67 proliferation index in breast cancer as a function of assessment method. A NordiQC experience. Appl Immunohistochem Mol Morphol. 2020;10.1097/PAI.0000000000000846.
doi:10.1097/PAI.0000000000000846. Online ahead of print.

Study 2.
Røge R, Riber-Hansen R, Nielsen S, Vyberg M. Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining. Breast Cancer Res Treat. 2016 Jul;158(1):11-19. doi: 10.1007/s10549-016-3852-6.

Study 3.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Impact of Primary Antibody Clone, Format, and Stainer Platform on Ki67 Proliferation Indices in Breast Carcinomas. Appl Immunohistochem Mol Morphol. 2019 Nov/Dec;27(10):732-739.

Study 4.
Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Potential of Cell Lines Assessed by Image Analysis Used as Performance Controls of Ki67 Immunostained. Appl Immunohistochem Mol Morphol. 2020 Mar 9. doi: 10.1097/PAI.0000000000000845. Online ahead of print.