Virchows Arch (2014) 465:509–519
DOI 10.1007/s00428-014-1624-4

ORIGINAL ARTICLE

Evaluation of colon cancer histomorphology: a comparison between formalin and PAXgene tissue fixation by an international ring trial

Sibylle Gündisch · Julia Slotta-Huspenina · Paolo Verderio · Chiara Maura Ciniselli · Sara Pizzamiglio · Christina Schott · Enken Drecoll · Christian Viertler · Kurt Zatloukal · Marcel Kap · Peter Riegman · Irene Esposito · Katja Specht · Gregor Babaryka · Martin Asslaber · Koppany Bodó · Michael den Bakker · Jan den Hollander · Falko Fend · Jens Neumann · Simone Reu · Aurel Perren · Rupert Langer · Alessandro Lugli · Ingrid Becker · Thomas Richter · Gian Kayser · Annette M. May · Fatima Carneiro · José Manuel Lopes · Leslie Sobin · Heinz Höfler · Karl-Friedrich Becker

Received: 10 April 2014 / Revised: 18 June 2014 / Accepted: 30 June 2014 / Published online: 2 August 2014

Abstract The aim of our study was to evaluate the quality of histo- and cytomorphological features of PAXgene-fixed specimens and their suitability for histomorphological classification in comparison to standard formalin fixation. Fifteen colon cancer tissues were collected, divided into two mirrored samples and either formalin fixed (FFPE) or

Sibylle Gündisch and Julia Slotta-Huspenina contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00428-014-1624-4) contains supplementary material, which is available to authorized users.

S. Gündisch · J. Slotta-Huspenina · C. Schott · E. Drecoll · I. Esposito · K. Specht · G. Babaryka · H. Höfler · K.-F. Becker
Institute of Pathology, Technische Universität München, Trogerstr. 18, 81675 Munich, Germany

P. Verderio · C. M. Ciniselli · S. Pizzamiglio
Unit of Medical Statistics, Biometry and Bioinformatics, Fondazione IRCCS, Istituto Nazionale dei Tumori, via Venezian 1, 20133 Milan, Italy

C. Viertler · K. Zatloukal · M. Asslaber · K. Bodó
Institute of Pathology, Medical University of Graz, Auenbruggerplatz 25, 8036 Graz, Austria

M. Kap · P. Riegman · M. den Bakker · J. den Hollander
Department of Pathology, Erasmus Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

F. Fend
Institute of Pathology, Eberhard-Karls-University, Liebermeisterstr. 8, 72076 Tübingen, Germany

J. Neumann · S. Reu
Institute of Pathology, Ludwig-Maximilians-Universität München, Thalkirchner Str. 36, 80337 Munich, Germany

A. Perren · R. Langer · A. Lugli
Institute of Pathology, Universität Bern, Murtenstr. 31, Postfach 62, 3010 Bern, Switzerland

I. Becker · T. Richter
Pathology Rosenheim, Lilienweg 12, 83022 Rosenheim, Germany

G. Kayser · A. M. May
Department of Pathology, Ludwig-Aschoff-Haus, University Medical Center Freiburg, Breisacher Str. 115a, 79106 Freiburg, Germany

F. Carneiro · J. M. Lopes
IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

L. Sobin
Frederick National Laboratory for Cancer Research, The Cancer Human Biobank, National Cancer Institute, 6110 Executive Blvd, Suite 250, Rockville, MD 20852, USA

S. Gündisch · P. Verderio · C. M. Ciniselli · S. Pizzamiglio · C. Schott · C. Viertler · K. Zatloukal · M. Kap · P. Riegman · K.-F. Becker (✉)
The SPIDIA Consortium
URL: www.spidia.eu
e-mail: kf.becker@lrz.tum.de

PAXgene fixed (PFPE) before paraffin embedding. HE- and PAS-stained sections were scanned and evaluated in a blinded, randomised ring trial by 20 pathologists from Europe and the USA using virtual microscopy. The pathologists evaluated histological grading, histological subtype, presence of adenoma, presence of lymphovascular invasion, quality of histomorphology and quality of nuclear features. Statistical analysis revealed that the reproducibility with regard to grading between both fixation methods was rather satisfactory (weighted kappa statistic (k_w)=0.73 (95 % confidence interval (CI), 0.41–0.94)), with a higher agreement between the reference evaluation and the PFPE samples (k_w=0.86 (95 % CI, 0.67–1.00)). Independent from preservation method, inter-observer reproducibility was not completely satisfactory (k_w=0.60). Histomorphological quality parameters were scored equal or better for PFPE than for FFPE samples. For example, overall quality and nuclear features, especially the detection of mitosis, were judged significantly better for PFPE cases. By contrast, significant retraction artefacts were observed more frequently in PFPE samples. In conclusion, our findings suggest that the PAXgene Tissue System leads to excellent preservation of histomorphology and nuclear features of colon cancer tissue and allows routine morphological diagnosis.

Keywords Histomorphology · Molecular diagnostic · Colon cancer · Tissue preservation · Formalin free · Reproducibility

Introduction

In routine clinical practice, the gold standard for histomorphological analysis is formalin-fixed, paraffin-embedded (FFPE) tissues. However, there are two major reasons why formalin is not the ideal preservation method. First, it is recognised that the cross-linking property of formaldehyde is problematic for many downstream molecular analyses and second, the carcinogenic capacity of formaldehyde holds a serious environmental risk [3, 21]. An alternative are cryopreserved tissue specimens which offer the possibility for sophisticated molecular analysis, but the histomorphology is less well preserved and handling during clinical procedures is more complicated. Thus, there is a need for novel fixation solutions which allow both high-quality molecular and histopathological analyses.

Within the large-scale European project *Standardisation and Improvement of Generic Pre-analytical Tools and Procedures for In Vitro Diagnostics* (SPIDIA) and other research projects, it has previously been shown that the novel formalin-free fixation technology PAXgene Tissue System simultaneously preserves tissue morphology and antigenicity as well as nucleic acids, proteins and phosphoproteins in clinical tissue samples [8, 10, 11, 17, 23, 26, 28]. Compared with formalin the PAXgene Tissue System is a non-cross-linking, non-carcinogenic, mixture of different alcohols, acid and a soluble organic compound that rapidly preserves morphology as well as all biomolecules. In order to improve acceptance in routine clinical applications, especially for histopathological cancer diagnostics compared with FFPE, the suitability of PAXgene tissue still needs to be shown by independent and blinded ring trials.

Thus, the aim of our study was to evaluate the quality of histo- and cytomorphological features of PAXgene-fixed tissues and the suitability of the PAXgene tissue preservation technology for histomorphological classification of colon cancer in comparison to the state-of-the-art FFPE technique. Because all previous studies have been performed mainly on non-malignant tissue samples, we decided to focus on colon cancer, which is the fourth most common malignancy in the Western world and the third most frequent cause of cancer-related mortality [15, 16]. We applied virtual microscopy to ensure that the participants evaluated the same section with the same cellular components. Different studies have previously shown that diagnostic performance of virtual slides is at least equal, if not superior compared with conventional microscopy [18, 25]. Comprehensive statistical analysis was performed to evaluate the potential of the PAXgene Tissue System for routine diagnostics of colon cancer.

Materials and methods

Tissue samples

Tissue samples of 15 colon cancer cases (13 adenocarcinomas not otherwise specified (NOS) and two mucinous adenocarcinomas) were prospectively collected at the Institute of Pathology of the Technische Universität München, Munich, Germany in close collaboration with surgeons from the affiliated Department of Surgery. Samples were obtained from patients diagnosed with colon carcinoma if enough residual tumour tissue not needed for routine diagnostics was left for research purposes. All patients signed written informed consent, and the study was approved by the Ethics Committee of the Klinikum Rechts der Isar of the Technische Universität München (reference number 2336/09). Immediately after surgery, each tumour specimen was divided into two mirrored aliquots and either fixed in 10 % neutral-buffered formalin or fixed with PAXgene Tissue Fix (PreAnalytix GmbH, Hombrechtikon, Switzerland) for 24 h. One case was fixed for 72 h and two cases for 96 h, both in formalin and PAXgene Tissue Fix. Subsequently, PAXgene-fixed samples were transferred into PAXgene Tissue Stabiliser to stop the fixation process. Using a standard protocol, PAXgene-treated and formalin-fixed tissues were dehydrated and embedded in low-melting temperature paraffin in two different embedding

processors to avoid cross-contamination of the two fixatives. Sections from all samples were stained with haematoxylin and eosin (HE) and periodic acid-Schiff (PAS) and digitised using the virtual microscopy platform of the Erasmus Medical Center, Rotterdam, The Netherlands (Fig. 1). Slides were scanned at ×40 resolution, comparable to ×400 magnification of conventional light microscopy, using a Nanozoomer Digital Pathology (NDP) slide scanner (Hamamatsu, Japan). Digitised slides were coded according to the randomization scheme provided by the project statisticians (P.V., C.C. and S.P.) and uploaded to a password protected internet folder. A short user manual for the NDP viewer software and morphometric features therein was provided with the study protocol that was sent to each participant. Reference values were determined by independent evaluation of the "gold standard" FFPE virtual slides by two pathologists of the Technische Universität München (J. S.-H. and E. D.) according to the WHO classification criteria (>95 % gland formation, G1, well differentiated; 50–95 % gland formation, G2, moderately differentiated; 0–49 % gland formation, G3, poorly differentiated) [4] and in case of disagreement consensus was reached by joint examination. The pathologists paid attention that all features to be evaluated were similarly present in both fixed samples of each case. Lymphovascular invasion (LVI) was defined as present if invasion could be detected in the HE-stained FFPE virtual slide and has been proven by immunohistochemistry against the markers CD34 and D2-40 [12, 19, 22] (Online resource 1).

Study design

Twenty pathologists from five European countries (Austria, Germany, Portugal, Switzerland and The Netherlands) and the USA agreed to participate in the ring trial. Links to the randomised virtual slides were distributed to the participants together with the study protocol, the evaluation form, the detailed instructions and the timeline for the ring trial implementation. Each

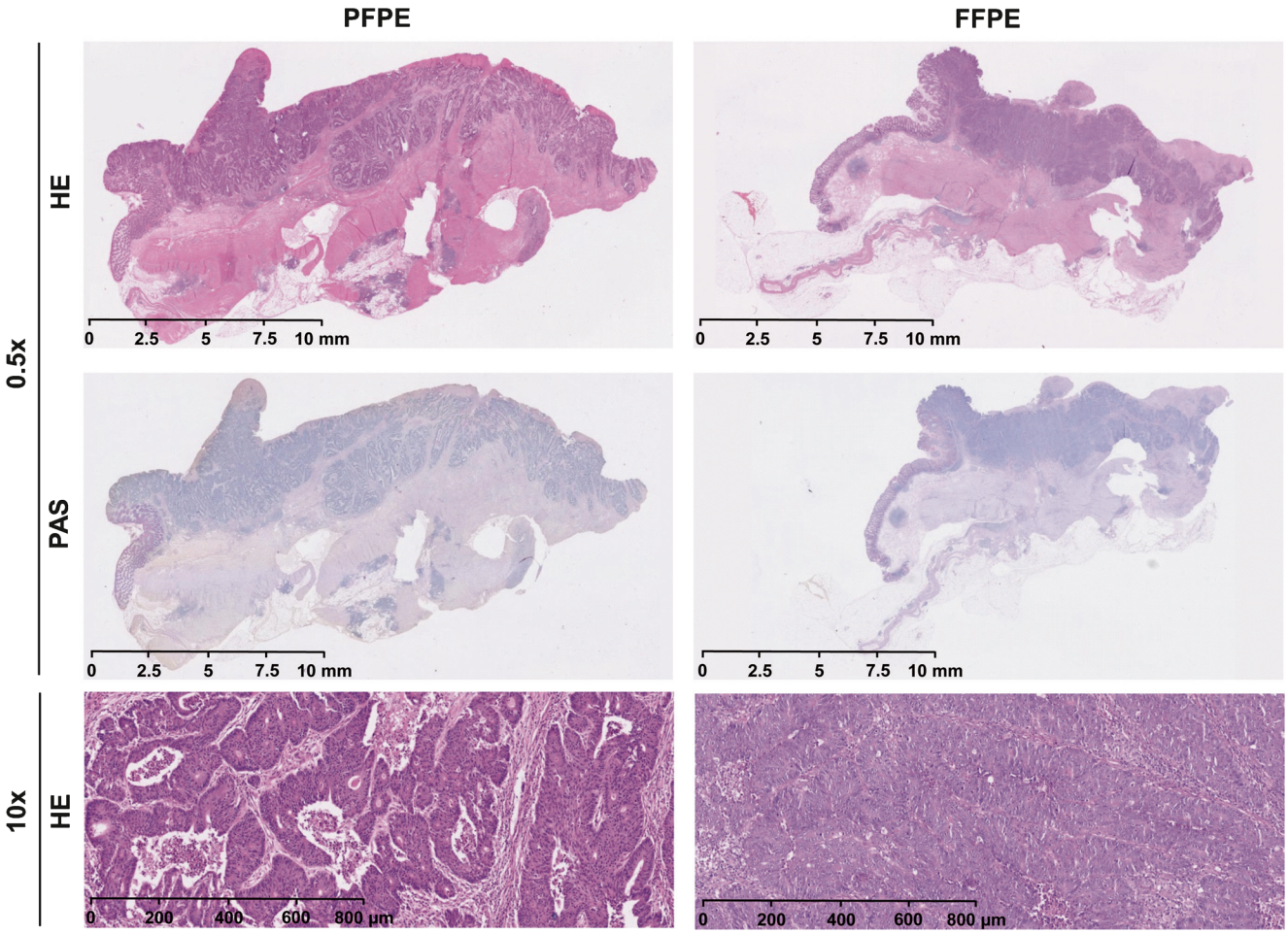


Fig. 1 Overview picture of one mirrored sample. Haematoxylin and eosin (*HE*) and periodic acid-Schiff (*PAS*) staining of one exemplary mirrored sample (case ID 1), which had been either PAXgene fixed and paraffin embedded (*PFPE*) or formalin fixed and paraffin embedded (*FFPE*). Original magnification, ×0.5 and ×10

participant evaluated only one of the two virtual slides obtained from each selected case, either the FFPE or the PFPE sample, for which the fixation method was blinded. The selection of the cases was performed according to a randomised scheme implying the subdivision of both the differently fixed cases and participants in two groups (Fig. 2). In particular, each participant evaluated histological grading according to the WHO classification criteria (percentage of glandular formation), histological subtype, presence of adenoma, presence of LVI, quality of histomorphology and quality of nuclear features.

Statistical analysis

The reproducibility between fixation methods (PFPE vs. FFPE) and the reference scenario (RS), as well as inter-observer reproducibility was assessed by considering the histological grading (classified as G1, G2 and G3/G4) score as the pivotal variable.

Due to the nature of the grading score, the reproducibility was evaluated by computing the weighted kappa statistic (k_w) [9]. This statistic is the most widely accepted measure of agreement if, as in our case, the data in question arise from an ordinal scale. Its values are usually between 0 (absence of agreement) and 1 (absolute agreement). A negative value may be obtained in situations where the actual agreement is less than a chance one. As previously reported [24], the observed values of k_w were considered satisfactory if equal to or greater than 0.80. For the reproducibility between fixation methods as well as between reference evaluation and fixation method, all the participants' data were used to compute the modal category within each method for each case (modal scenario (MS)). Then, by starting from these modal values, the jackknifed estimate of the weighted kappa statistic (k_{wj}) was computed together with the relative 95 % confidence interval (CI) [7]. In addition, for the inter-observer reproducibility the kappa category-specific statistics (k_{cs}) and their weighted averages (Cohen's kappa statistic (k_c)) were estimated by jointly considering all participants data [9, 13]. Each k_{cs} value was interpreted in a qualitative manner on the basis of the Landis and Koch classification criteria [20].

The strength of association between fixation methods and the pivotal variables (presence of adenoma, histological subtype, presence of LVI and general quality parameters) was assessed by the Fisher exact test [1] by taking advantage of the MS. Cases showing bimodal distribution were excluded from each specific analysis. All statistical analyses were performed with the SAS

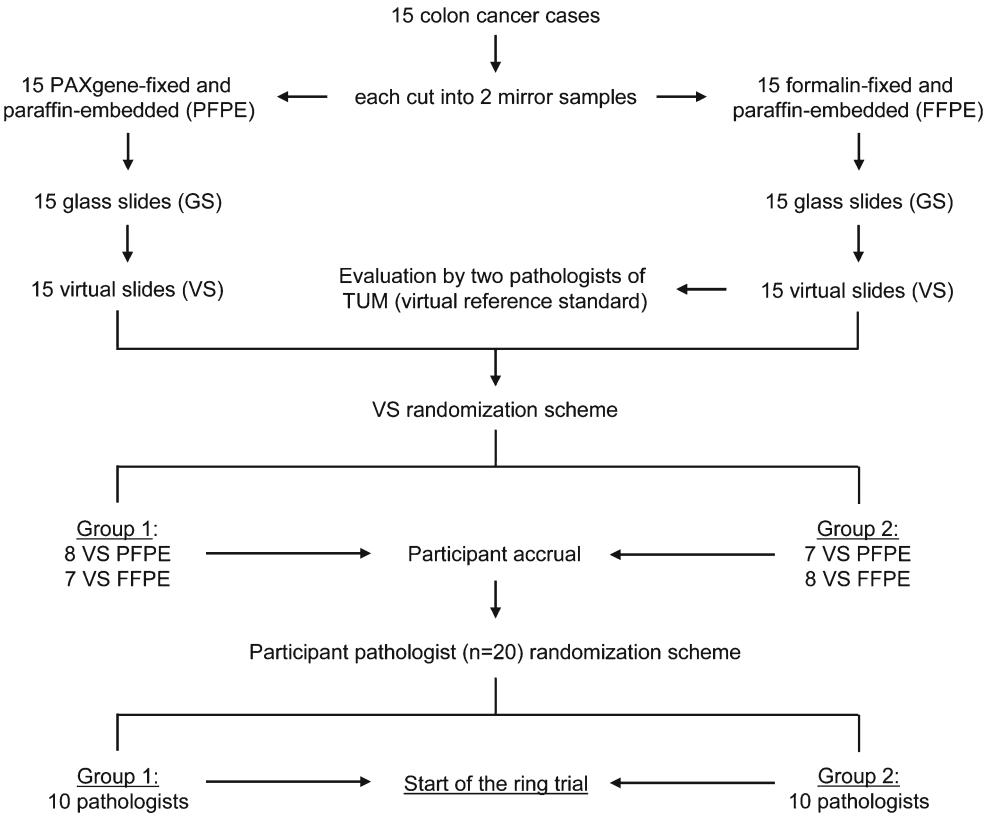


Fig. 2 Study design. Each participant evaluated only one of the two virtual slides obtained from each selected case, either the FFPE or the PFPE sample, and the fixation method was blinded. The selection of the cases was performed according to a randomised scheme implying the subdivision of both, the differently fixed cases and the participants in two groups

software (version 9.2; SAS Institute Inc. Cary, NC) by adopting a significance alpha level of 0.05.

Results

Reproducibility of colon cancer grading between the fixation methods and vs. the reference scenario

The reproducibility of colon cancer grading according to WHO classification criteria between both fixation methods was rather satisfactory with a k_{wj} value of 0.73 (95 % CI, 0.41–0.94). The concordance table depicted in Table 1 shows that four cases (three FFPE and one PFPE) had to be excluded from statistical analysis because of a bimodal distribution of the grading, each being classified by five pathologists as G2 and by five pathologists as G3/G4. Interestingly, the corresponding PFPE samples of those three excluded FFPE cases showed a concordant classification compared to the reference value whereas vice versa this was not the case. Three of the remaining 11 cases resulted in a different grading for PFPE and FFPE. Interestingly, for all three cases, the FFPE sample was graded higher compared with the corresponding PFPE sample. In detail, two cases were graded as G1 in the PFPE samples, concordant to the reference value, but were categorised as G2 in the FFPE samples. Additionally, one case was graded as G2 in the PFPE sample, concordant to the reference evaluation, but defined as G3/G4 in the FFPE sample.

Thus, the reproducibility of colon cancer grading between the RS and MS according to the respective fixation method was rather satisfactory for FFPE with a k_{wj} value of 0.62 (95 % CI, 0.29–0.83) and satisfactory for PFPE with a k_{wj} value of 0.86 (95 % CI, 0.67–1.00). The concordance tables depicted in Table 2 show that five FFPE cases were classified differently compared with the reference value, four of which showed a higher grading. By contrast, only two PFPE cases showed a different classification compared with the reference value, only one of which was graded higher.

Table 1 Reproducibility between fixation methods

PFPE	FFPE			
	G1	G2	G3/G4	Total
G1	1	2	0	3
G2	0	4	1	5
G3/G4	0	0	3	3
Total	1	6	4	11

Table 2 Reproducibility between reference and modal scenario within fixation method

	G1	G2	G3/G4	Total
RS	MS FFPE			
G1	1	3	0	4
G2	0	2	1	3
G3/G4	0	1	3	4
Total	1	6	4	11
RS	MS PFPE			
G1	3	1	0	4
G2	0	3	0	3
G3/G4	0	1	3	4
Total	3	5	3	11

RS reference scenario, *MS* modal scenario

Inter-observer reproducibility of colon cancer grading

Twenty pathologists (ten for group 1 and ten for group 2) scored the virtual slides for colon cancer grading according to the WHO classification criteria. Among these, four participants did not provide the grading score for each case and thus could not be considered in the analysis of the inter-observer reproducibility. One reason for missing grading values in the evaluation sheets was the presence of two mucinous adenocarcinomas (>50 % extracellular mucin) which were not graded by each pathologist and a case which was defined as a medullary carcinoma by only one pathologist. In general, the level of agreement within each group was not completely satisfactory, with a median k_w value of 0.60 (range, 0.32–1.00) and 0.60 (range, 0.41–0.91) for groups 1 and 2, respectively. The distribution of the k_w values of each participant compared with all the others, as well as for the overall series, are displayed by a specific box-plot in Fig. 3. The interchangeability of the two groups was confirmed by the superimposable level of agreement observed by considering the k_w distribution in the overall series within each group. Comparison of reference grading values and each group resulted in an even lower median k_w value of 0.54 (range, 0.17–0.71) and 0.56 (range, 0.43–0.82) for groups 1 and 2, respectively. Only one pathologist showed a satisfactory agreement (k_w=0.82) compared with the reference evaluation, in contrast the lowest k_w value was 0.17.

As shown in Table 3, in both groups the most important contribution to the overall agreement (corresponding to a k_c of 0.46 and 0.44 for groups 1 and 2, respectively) is related to the category G3/G4 with a k_{cs} of 0.53 for group 1 and 0.54 for group 2. By contrast, the grading category G2 appeared to provide the poorest contribution to the observed overall agreement with a k_{cs} of 0.39 for group 1 and 0.30 for group 2.

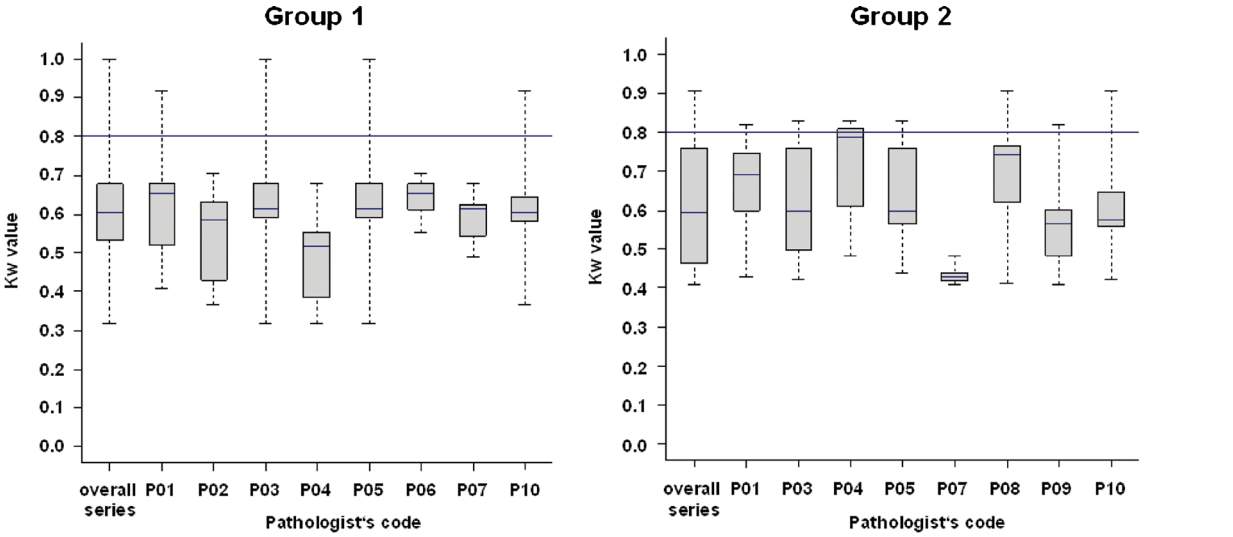


Fig. 3 Inter-observer reproducibility. The distribution of the k_w values of each participant compared with all the others, as well as for the overall series, is displayed within each group by a specific box-plot. In each box-plot, the *two horizontal sides of the box* identify the 25th and 75th percentiles; the *horizontal line inside the box* indicates the median and *the limits of the two whiskers* correspond to minimum and maximum of the overall distribution. The *horizontal line* represents the threshold value of k_w=0.80. The pathologists P08 and P09 from group 1 and P02 and P06 from group 2 were not considered in this analysis because they did not provide the grading score for each case

Other considered parameters

Besides histological grading of colon cancer additional parameters, namely the presence of adenoma, the histological subtype and the presence of LVI, were evaluated with regard to the fixation method. We could not detect significant associations between those parameters and the respective fixation method which is depicted in Tables 4, 5 and 6. The most critical issue in this analysis was the evaluation of LVI. Some pathologists commented on this issue that it is impossible to distinguish between LVI and retraction artefacts, nevertheless, everybody evaluated it. Interestingly, even though the retraction artefacts were more prominent in PFPE samples the evaluation of LVI in those samples showed a higher concordance to the RS than the FFPE samples. We detected three strong misclassifications (present vs. not seen) in the FFPE samples whereas only one for the PFPE samples. Furthermore, by Fisher exact test we could not find a statistically significant association between retraction artefacts and the presence of LVI, as well as the presence of adenoma or histological subtype (data not shown).

Table 3 Inter-observer reproducibility

	G1	G2	G3/G4
k_{cs} value for group 1			
G1	0.49	–	–
G2	–	0.39	–
G3/G4	–	–	0.53
Cohen's k statistic (k_c), 0.462			
k_{cs} value for group 2			
G1	0.51	–	–
G2	–	0.30	–
G3/G4	–	–	0.54
Cohen's k statistic (k_c), 0.436			

Association between quality parameters and respective fixation method

A total of 14 quality parameters for histomorphology (preservation of epithelium, mucus in HE stain, mucus in PAS stain, basal lamina, fatty tissue, tunica muscularis, necrosis, retraction artefacts, edge artefacts and overall quality) and nuclear features (chromatin, nucleoli, mitosis and apoptosis) were investigated with respect to the fixation method. For three parameters, we found a statistically significant association as reported in Tables 7, 8 and 9. In these cases, the modal distribution was

Table 4 Association between presence of adenoma and fixation method

Adenoma	FFPE	PFPE
Not present	9	10
Present	5	4
Total	14	14

Bimodal cases: No. 3 FFPE—5 not present, 5 present

Table 5 Association between histological subtype and fixation method

Invasive adenocarcinoma	FFPE	PFPE
No special subtype	13	14
Mucinous adenocarcinoma (>50 % extra cellular mucin)	2	1
Signet-ring cell carcinoma (>50 % signet-ring cells)	0	0
Total	15	15

categorised according to the original classification (retraction artefacts) or according to a three-class reclassification (mucus in HE stain and overall quality). The evaluation of the quality of mucus preservation in the HE staining (p value=0.014) and the overall quality (p value=0.016) revealed that no case was judged as poor/weak in both fixation methods but all PFPE cases were classified as good/excellent, whereas six FFPE cases were classified as only satisfactory. The evaluation of retraction artefacts revealed that the majority of FFPE cases were classified with retraction artefacts not present or minor and none with moderate or significant retraction artefacts (p value=0.006). By contrast, in PFPE samples retraction artefacts were observed more frequently and the majority was evaluated with minor or moderate retraction artefacts present. Figure 4 depicts two examples of the virtual slides used in the ring trial showing HE- and PAS-stained PFPE and FFPE colon cancer tissue samples. The pictures clearly show that the staining for both HE, and also the PAS staining, were more intense in the PFPE tissues than in corresponding FFPE tissues. Some participants mentioned that the contrast in the HE-stained slides of PFPE samples was very good. This led for example to the fact that nuclear features like mitosis were also detectable more accurately; however, it was not statistically significant (p value=0.09).

Further general comments not related to the fixation method basically referred to the quality of the scan (e.g. some areas were out of focus) and the quality of the sections and staining (e.g. uneven staining, wrinkles, tears, floaters, poor contrast and section too thick).

Table 6 Association between presence of lymphovascular invasion and fixation method

Lymphovascular invasion	FFPE	PFPE
Present (L1/V1)	2	5
Possible	1	1
Not seen (L0/V0)	9	6
Total	12	12

Bimodal cases: No. 6 PFPE—4 possible, 4 not seen; No. 7 PFPE—4 present, 4 not seen; No. 10 PFPE—5 possible, 5 not seen

Table 7 Association between quality of mucus preservation (HE staining) and fixation method

Mucus HE	FFPE	PFPE
Poor/weak	0	0
Satisfactory	6	0
Good/excellent	6	12
Total	12	12

Fisher exact text, p value=0.014. Bimodal cases: No. 2 FFPE—4 satisfactory, 4 good; No. 2 PFPE—4 satisfactory, 4 good; No. 4 FFPE—4 satisfactory, 4 good; No. 4 PFPE—4 satisfactory, 4 good; No. 7 PFPE—4 satisfactory, 4 good

Discussion and conclusion

In this study, we report the results of a systematic comparative evaluation of colon cancer histomorphology between conventional formalin fixation and novel PAXgene tissue fixation, performed by 20 pathologists from Europe and the USA.

The suitability of the novel preservation technology PAXgene Tissue System for colon cancer grading was evaluated by comparing the results from mirrored FFPE or PAXgene-fixed or PFPE clinical tissue samples. The reproducibility between both fixation methods using the MS was rather satisfactory (k_w=0.73). Three cases resulted in misclassification between FFPE and PFPE, and it was remarkable that all FFPE cases were graded higher compared with the corresponding PFPE sample, which showed the same grading as the RS. The overall evaluation of the reproducibility of colon cancer grading between the RS and the MS concerning the respective fixation method confirmed this finding. There was higher agreement in grading between the PFPE samples and the reference evaluation (k_w=0.86) than for FFPE samples (k_w=0.62). This was an interesting finding because our study reference was determined by independent evaluation of the FFPE virtual slides by two pathologists from one institute (Institute of Pathology of the Technical University Munich). Thus, one would have expected that the FFPE samples would

Table 8 Association between retraction artefacts and fixation method

Retraction artefacts	FFPE	PFPE
Not present	8	1
Minor	2	6
Moderate	0	2
Significant	0	1
Total	10	10

Fisher exact text, p value=0.006. Bimodal cases: No. 4 FFPE—3 not present, 3 minor, 3 moderate (trimodal); No. 5 PFPE—4 minor, 4 moderate; No. 6 FFPE—5 not present, 5 minor; No. 9 FFPE—4 minor, 4 moderate; No. 11 FFPE—3 not present, 3 minor

Table 9 Association between overall quality of morphology and fixation

Overall quality	FFPE	PFPE
Poor/weak	0	0
Satisfactory	6	0
Good/excellent	8	14
Total	14	14

Fisher exact text, p value=0.016. Bimodal cases: No. 2 FFPE—5 satisfactory, 5 good

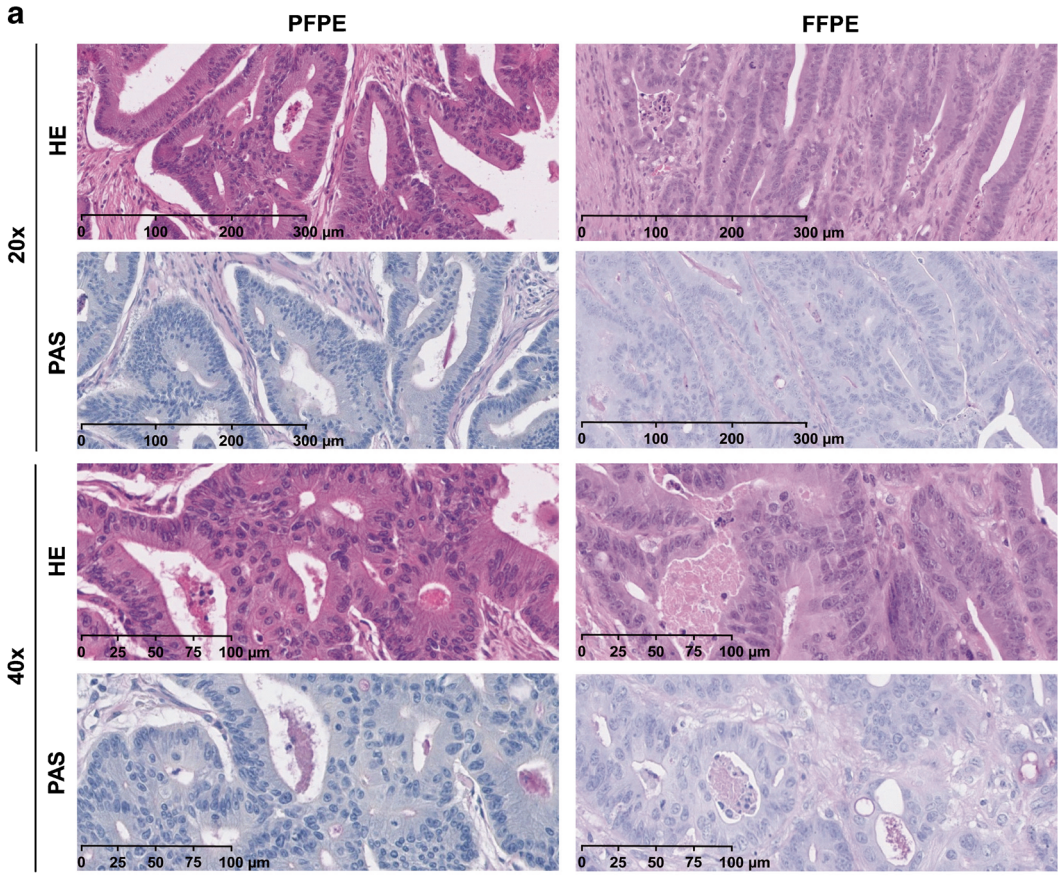
show greater agreement with the reference values. A reason for these diverging results might be that in the reference evaluation the percentage of gland formation was analysed strictly according to the WHO classification criteria without consideration of cytomorphology. In the PFPE samples, the glands could be detected more accurately (Figs. 1 and 4) allowing the participants to focus mainly on the percentage of gland formation which resulted in a high agreement between PFPE samples and the RS. By contrast, it seems that the assessment of tumour differentiation of FFPE samples was often influenced by cytomorphological features instead of gland formation only, which might have led to higher grading in the end.

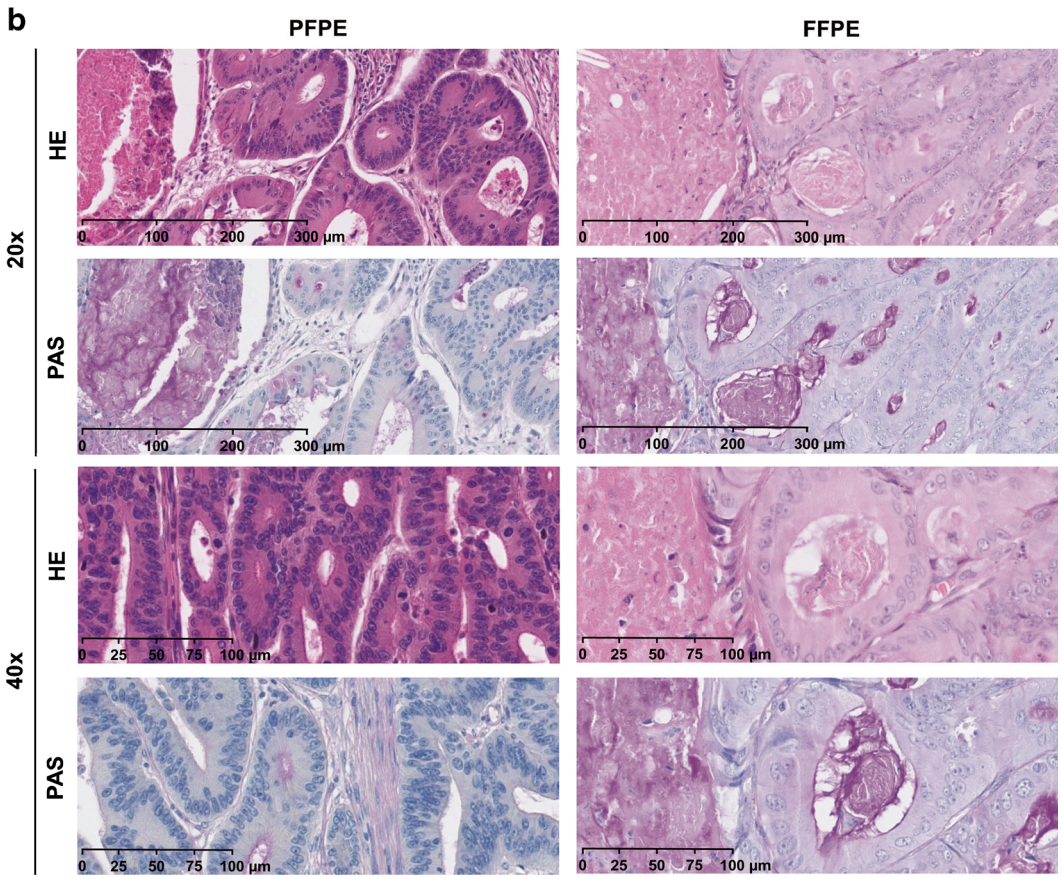
Inter-observer reproducibility concerning histological grading within the two groups of pathologists was not completely satisfactory (k_w=0.60). This finding confirms previous studies reporting that assessment of tumour differentiation in general shows high inter-observer variability [5, 6, 27]. The level of agreement concerning histopathological grading of colorectal cancer in these studies was similar or even lower compared to our ring trial. Furthermore, the performance level in evaluating the grading seemed to be similar in the two considered groups of pathologists, which confirmed the appropriateness of our randomization scheme. The reproducibility of histological grading between the ring trial results and the results of the RS was even slightly lower (k_w=0.54 and 0.56 for groups 1 and 2, respectively) but again nearly no differences could be detected between the two groups of pathologists. The most important contribution to the observed overall agreement was related to the grading category G3/G4, and the most critical grading category in this context was G2. It has been already shown in other studies that high inter-observer variability mainly concerns grade 2 tumours [2, 14]. The major problem seems the subjective nature of assessment of tumour differentiation (percentage of gland formation) and the lack of possibilities for objective quantification of specific parameters [18]. Interestingly, those pathologists showing the highest agreement (k_w>0.8) within a group were not from the same department, where one would have expected common training but from different countries.

In a previous study, assessing histomorphology of PAXgene-fixed tissues artefacts such as cell shrinkage/retraction was occasionally observed [17]. Thus, we aimed to clarify within an international ring trial whether these artefacts may influence a proper diagnosis in tumour tissues. Therefore, besides histological grading of the colon cancer samples, additional parameters were evaluated with regard to the fixation method. No significant associations could be detected between presence of adenoma, histological subtype of carcinoma and presence of LVI. Although, there was no statistical relevance the most critical issue in this context was the evaluation of LVI in the presence of retraction artefacts. Our study revealed that retraction artefacts were significantly more frequent in PFPE samples (p value=0.006) but interestingly the concordance in evaluating presence of LVI in PFPE samples compared with the RS was higher than for FFPE samples. These data suggest that the evaluation of LVI in the presence of retraction artefacts is a common problem not related to the fixation method. It was already published for FFPE samples that standard pathological methods like HE staining are not sufficient for direct observation of tumour cell infiltration into the vessels and that evaluation of LVI is complicated by tissue retraction artefacts and high inter-observer variability [12, 19, 22]. Thus, it is generally recommended to perform immunohistochemical analysis using markers for vascular and lymphatic channels to validate invasion [19, 22]. We also performed immunohistochemistry against CD34 and D2-40 with our reference slides to prove LVI. For future approaches, we would recommend immunohistochemistry also for PFPE samples to be able to reliably differentiate LVI from retraction artefacts.

Concerning the quality parameters for histomorphology and nuclear features, three out of 14 parameters showed a statistically significant association with the fixation method. Specifically, the overall quality of histomorphology and mucus preservation was judged significantly better for PFPE cases (p value=0.016 and 0.014, respectively). Other considered parameters were not significantly related to the fixation method. Interestingly, the HE and PAS staining was more intense in PFPE samples, and it was mentioned by some participants that the contrast in those HE-stained slides was very good. This led for example to the fact that nuclear features like detection of mitosis were also scored equal or better for PFPE samples, however it was not statistically significant (p value=0.09). This might also facilitate accurate grading in cancer tissues and could be an explanation for high agreement in grading between PFPE samples and the RS.

Fig. 4 Examples for overall morphology of PFPE and FFPE tissue samples. Haematoxylin and eosin (*HE*) and periodic acid-Schiff (*PAS*) staining of two examples (**a**, case ID 10; **b** case ID 8) for G1 colorectal cancers which had been either PAXgene fixed and paraffin embedded (*PFPE*) or formalin fixed and paraffin embedded (*FFPE*). Original magnification, ×20 and ×40


a
PFPE
FFPE
20x
40x
HE
PAS
0 100 200 300 µm
0 25 50 75 100 µm


b
PFPE
FFPE
20x
40x
HE
PAS
0 100 200 300 µm
0 25 50 75 100 µm

Further comments not related to the fixation method basically referred to the quality of the sections and staining's and the quality of the scan. Although the use of virtual microscopy was in general well accepted by the participants some of them mentioned that a few areas of the scanned slides were out of focus. Nevertheless, the diagnostic reproducibility might not be affected by this issue because the reference evaluation was performed with the same virtual slides.

In parallel to our ring trial focussing on colon cancer, partners from the SPIDIA consortium performed similar studies with breast cancer (Viertler et al., in preparation) and prostate cancer samples (Kap et al., in preparation). Within these three different morphology ring trials we achieved participation of more than 70 renowned pathologists from Europe and the USA. In future studies, further tumour types, such as papillary thyroid carcinomas or neuroendocrine tumours have to be evaluated where careful attention to nuclear details and assessment of mitotic rate is required. In addition, large tissue specimens like whole hemicolectomy specimens should be assessed. Preliminary results obtained by fixation of pig organs suggest that when tissue is pretreated as usual, e.g. incision(s) to allow fixation, PAXgene is able to fix whole organs overnight [17]. So far, our findings indicate that the PAXgene Tissue System leads as compared with formalin fixation to similar or even better preservation of histomorphology including nuclear features and allows accurate morphological diagnosis. Thus, it has great potential to serve as a multimodal fixative for pathology in personalised medicine, enabling histopathological diagnosis and a broad spectrum of biomarker analyses from the same clinical tissue specimen.

Acknowledgments This work was performed within the European consortium Standardisation and Improvement of Generic Pre-analytical Tools and Procedures for In Vitro Diagnostics (SPIDIA; www.spidia.eu), which is funded by the European Union within the Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 222916.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Agresti A (ed) (1990) Categorical data analysis. John Wiley and Sons, New York
2. Barresi V, Reggiani Bonetti L, Branca G et al (2012) Colorectal carcinoma grading by quantifying poorly differentiated cell clusters is more reproducible and provides more robust prognostic information than conventional grading. Virchows Arch 461:621–628
3. Bosetti C, Mclaughlin JK, Tarone RE et al (2008) Formaldehyde and cancer risk: a quantitative review of cohort studies through 2006. Ann Oncol Off J Eur Soc Med Oncol ESMO 19:29–43
4. Bosman FT, Carneiro F, Hruban RH et al (eds) (2010) WHO classification of tumours of the digestive system. International Agency for Research on Cancer (I A R C)
5. Chandler I, Houlston RS (2008) Interobserver agreement in grading of colorectal cancers—findings from a nationwide Web-based survey of histopathologists. Histopathology 52:494–499
6. Chetty R, Gill P, Govender D et al (2012) International study group on rectal cancer regression grading: interobserver variability with commonly used regression grading systems. Hum Pathol 43:1917–1923
7. Corletto V, Verderio P, Giardini R et al (1998) Evaluation of residual cellularity and proliferation on preoperatively treated breast cancer: a comparison between image analysis and light microscopy analysis. Anal Cell Pathol 16:83–93
8. Ergin B, Meding S, Langer R et al (2010) Proteomic analysis of PAXgene-fixed tissues. J Proteome Res 9:5188–5196
9. Fleiss JL, Levin B, Paik MC (eds) (1981) Statistical methods for rates and proportions. John Wiley and Sons, New York
10. Groelz D, Sobin L, Branton P et al (2013) Non-formalin fixative versus formalin-fixed tissue: a comparison of histology and RNA quality. Exp Mol Pathol 94:188–194
11. Gündisch S, Schott C, Wolff C et al (2013) The PAXgene((R)) tissue system preserves phosphoproteins in human tissue specimens and enables comprehensive protein biomarker research. PLoS One 8: e60638
12. Harris EI, Lewin DN, Wang HL et al (2008) Lymphovascular invasion in colorectal cancer: an interobserver variability study. Am J Surg Pathol 32:1816–1821
13. Holman CD (1984) Analysis of interobserver variation on a programmable calculator. Am J Epidemiol 120:154–160
14. Italian Network for Quality Assurance of Tumour Biomarkers G (2005) Quality control for histological grading in breast cancer: an Italian experience. Pathologica 97:1–6
15. Jemal A, Center MM, Desantis C et al (2010) Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 19:1893–1907
16. Jemal A, Siegel R, Xu J et al (2010) Cancer statistics, 2010. CA Cancer J Clin 60:277–300
17. Kap M, Smedts F, Oosterhuis W et al (2011) Histological assessment of PAXgene tissue fixation and stabilization reagents. PLoS One 6: e27704
18. Kayser K (2012) Introduction of virtual microscopy in routine surgical pathology—a hypothesis and personal view from Europe. Diagn Pathol 7:48
19. Kim S, Park HK, Jung HY et al (2013) ERG immunohistochemistry as an endothelial marker for assessing lymphovascular invasion. Korean J Pathol 47:355–364
20. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
21. Nielsen GD, Wolkoff P (2010) Cancer effects of formaldehyde: a proposal for an indoor air guideline value. Arch Toxicol 84:423–446
22. O'donnell RK, Feldman M, Mick R et al (2008) Immunohistochemical method identifies lymphovascular invasion in a majority of oral squamous cell carcinomas and discriminates between blood and lymphatic vessel invasion. J Histochem Cytochem 56:803–810
23. Oetjen J, Aichler M, Trede D et al (2013) MRI-compatible pipeline for three-dimensional MALDI imaging mass spectrometry using PAXgene fixation. J Proteomics 90:52–60
24. Paradiso A, Ellis IO, Zito FA et al (2009) Short- and long-term effects of a training session on pathologists' performance: the INQAT experience for histological grading in breast cancer. J Clin Pathol 62:279–281
25. Shaw EC, Hanby AM, Wheeler K et al (2012) Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study. J Clin Pathol 65:403–408
26. Staff S, Kujala P, Karhu R et al (2013) Preservation of nucleic acids and tissue morphology in paraffin-embedded clinical samples: comparison of five molecular fixatives. J Clin Pathol 66:807–810

27. Thomas GD, Dixon MF, Smeeton NC et al (1983) Observer variation in the histological grading of rectal carcinoma. J Clin Pathol 36:385–391

28. Viertler C, Groelz D, Gündisch S et al (2012) A new technology for stabilization of biomolecules in tissues for combined histological and molecular analyses. J Mol Diagn 14:458–466