npg

Atypical ductal hyperplasia: interobserver and intraobserver variability

Rohit K Jain*, Rutika Mehta*, Rosen Dimitrov, Lisbeth G Larsson, Paul M Musto, Kurt B Hodges, Thomas M Ulbright, Eyas M Hattab, Narasimhan Agaram, Muhammad T Idrees and Sunil Badve

Department of Pathology and Laboratory Science, I U School of Medicine, Indianapolis, IN, USA

Interobserver reproducibility in the diagnosis of benign intraductal proliferative lesions has been poor. The aims of the study were to investigate the inter- and intraobserver variability and the impact of the addition of an immunostain for high- and low-molecular weight keratins on the variability. Nine pathologists reviewed 81 cases of breast proliferative lesions in three stages and assigned each of the lesions to one of the following three diagnoses: usual ductal hyperplasia, atypical ductal hyperplasia and ductal carcinoma *in situ*. Hematoxylin and eosin slides and corresponding slides stained with ADH-5 cocktail (cytokeratins (CK) 5, 14. 7, 18 and p63) by immunohistochemistry were evaluated. Concordance was evaluated at each stage of the study. The interobserver agreement among the nine pathologists for diagnosing the 81 proliferative breast lesions was fair (κ-value = 0.34). The intraobserver κ-value ranged from 0.56 to 0.88 (moderate to strong). Complete agreement among nine pathologists was achieved in only nine (11%) cases, at least eight agreed in 20 (25%) cases and seven or more agreed in 38 (47%) cases. Following immunohistochemical stain, a significant improvement in the interobserver concordance (overall κ-value = 0.50) was observed (P = 0.015). There was a significant reduction in the total number of atypical ductal hyperplasia diagnosis made by nine pathologists after the use of ADH-5 immunostain. Atypical ductal hyperplasia still remains a diagnostic dilemma with wide variation in both inter- and intraobserver reproducibility among pathologists. The addition of an immunohistochemical stain led to a significant improvement in the concordance rate. More importantly, there was an 8% decrease in the number of lesions classified as atypical ductal hyperplasia in favor of usual hyperplasia; in clinical practice, this could lead to a decrease in the number of surgeries carried out for intraductal proliferative lesions.
Modern Pathology (2011) **24,** 917–923; doi:10.1038/modpathol.2011.66; published online 29 April 2011

Significant progress has been made in the diagnosis and treatment of breast cancer in the last few decades. Screening mammograms have made it possible to detect many tumors at an earlier stage and provide prompt treatment. Mammography has led to detection of increased numbers of breast lesions and subsequent diagnostic biopsies.[1] A spectrum of lesions from benign (usual ductal hyperplasia), borderline (atypical ductal hyperplasia), pre-invasive (ductal carcinoma *in situ*) to invasive (invasive ductal carcinoma)[2–5] is identified on these biopsies. As usual, ductal hyperplasia carries minimal or no increased risk of breast cancer; these patients do not undergo any additional procedures. On the other hand, atypical ductal hyperplasia and ductal carcinoma *in situ* progress to invasive carcinoma in nearly 4–5% and 8–10% of cases, respectively.[6] Patients with these lesions are advised excision with the addition of radiation for those with ductal carcinoma *in situ*. Although the clinical guidelines are well laid out, the histological differentiation between atypical ductal hyperplasia and ductal carcinoma *in situ* has been difficult. Several previous studies have shown that the concordance among pathologists in diagnosing atypical ductal hyperplasia especially is very poor, giving rise to potential misclassifications in treatment protocols.[7–11] Since atypical ductal hyperplasia and ductal carcinoma *in situ* comprise 10%[12]

Correspondence: Professor S Badve, MD, FRCPath, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, 350 West 11th Street, CPL 4050, Indianapolis, IN 46202, USA.
E-mail: sbadve@iupui.edu
*These authors contributed equally to the work.

Pathologists' concordance in diagnosis of proliferative breast lesions
918
RK Jain *et al*

and 15–20%,[13] respectively, of mammographically detected breast lesions, it becomes important to provide diagnostic aid to pathologists to recognize these lesions, resulting in better reproducibility.

In this study, we investigated the reproducibility in the interpretation of these intraductal proliferative breast lesions among university-based surgical pathologists. We also explored the observers' consistency in diagnosing these lesions and the impact of the addition of an immunohistochemical marker as a potential tool to differentiate these lesions and improve concordance rate.

Materials and methods
Design of the Study

After approval from the Institutional Review Board, nine pathologists from Department of Pathology and Laboratory Medicine of Indiana University participated in this study and classified 81 challenging cases of noninvasive breast lesions into one of the following categories: usual ductal hyperplasia, atypical ductal hyperplasia and ductal carcinoma *in situ*. Pathologists analyzed one hematoxylin and eosin (H&E) slide from each case in the first and second round (stages 1 and 2) and one H&E slide with the corresponding ADH-5 immunostain in the third round (stage 3).

Selection of Cases

A set of 81 H&E stained slides, each containing a challenging intraductal proliferative lesion, was selected by one of the authors (SB). In each slide, the representative ductal lesion was encircled and the pathologists were asked to evaluate only the tissue present in the circled area.

Immunohistochemical Assay

An immunohistochemical cocktail antibody, ADH-5, which is composed of CK5, 14, 7, 18 and p63 antibodies, was used to assist the analysis. Immunohistochemistry was performed on the unstained slides of these 81 cases. ADH-5 immunohistochemistry staining was performed as per the manufacturer's protocol. Briefly, after deparaffinization, $4 \mu m$ sections were exposed to antigen retrieval solution (citrate buffer (pH 6.0)) in a Dako PT module (Dako, Carpinteria, CA, USA). The slides were then incubated with ADH-5 antibody (IP-360; Biocare Medical, Concord, CA, USA) and the reaction was visualized using multiplex secondary reagent (IPSC5004), (IP DAB and IP fast Red; Biocare Medical). Counterstaining with hematoxylin was performed.

Circulation of the Slides

The H&E slides of 81 cases were labeled with code numbers and were circulated among the participating pathologists in batches of 40 and 41 cases. At the end of this first stage of the study, the results were collected from the pathologists. After a period of at least 1 week, the slides labeled with different code numbers were recirculated among the pathologists in two batches of 40 and 41 cases. At the end of this second stage, results were collected from the pathologists. After another interval of at least 1 week, 75 H&E cases, (in six cases, the immunohistochemical slides did not have the lesion) labeled with different code numbers along with the corresponding ADH-5 immunostain (third stage), were circulated in two batches of 37 and 38 cases. Each of the pathologists evaluated only the marked lesions on the same H&E slides in the first, second and third stages. Thus, discrepancies in the evaluations among the pathologists cannot be attributed to the dissimilarity of the lesions. There was neither precirculation of a training set of slides nor was any advice given concerning the interpretation of the cases (except for the Biocare Medical product literature on ADH-5).

Diagnostic Criteria

The participants were asked to apply criteria that they use in their daily practice for diagnosing the proliferative breast lesions. In the third stage of the study, the participants were asked to evaluate the H&E slides in combination with the ADH-5 immunostain.

Statistical Analyses

All statistical analyses were performed using SPSS version 17.0. A κ coefficient for multiple readers was used to evaluate the interobserver reproducibility.[14] This coefficient is a measurement of agreement, taking into account the amount of expected agreement due to chance. If the agreement is no better than expected by chance, the value of κ coefficient is zero; while in a case of perfect agreement, it is one. Agreement is considered poor, fair, moderate, good, or very good when κ coefficients range from <0.2, 0.2 to 0.39, 0.4 to 0.59, 0.6 to 0.79, or 0.8 to 1, respectively.[15] Differences in κ-values across different categories were tested using paired *t*-test or ANOVA as appropriate.

Results

In stage 1 of the study, complete agreement among nine pathologists was achieved in only nine (11%) cases: seven usual ductal hyperplasia and two ductal carcinoma *in situ*. At least eight agreed in 20 (25%) cases and seven or more agreed in 38 (47%) cases. The κ-values for all possible comparisons among the nine pathologists are shown in Table 1. In these comparisons, the κ-value ranged from a minimum of 0.15 (poor) to a maximum of 0.56 (moderate). The mean

Pathologists' concordance in diagnosis of proliferative breast lesions
RK Jain et al
919

Table 1 Concordance for all pairwise comparisons between nine pathologists for stage 1/stage 2/stage 3

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P2	0.57/0.74/0.74								
P3	0.28/0.25/0.62	0.23/0.33/0.61							
P4	0.40/0.61/0.76	0.33/0.44/0.71	0.33/0.16/0.73						
P5	0.40/0.56/0.36	0.25/0.50/0.31	0.20/0.26/0.33	0.37/0.50/0.39					
P6	0.37/0.40/0.45	0.32/0.30/0.52	0.32/0.24/0.43	0.50/0.48/0.47	0.56/0.34/0.28				
P7	0.40/0.44/0.59	0.40/0.29/0.62	0.15/0.26/0.48	0.56/0.48/0.62	0.32/0.42/0.30	0.39/0.40/0.54			
P8	0.47/0.30/0.83	0.25/0.28/0.70	0.23/0.33/0.62	0.21/0.29/0.78	0.37/0.34/0.35	0.28/0.21/0.37	0.43/0.46/0.55		
P9	0.36/0.57/0.54	0.24/0.42/0.47	0.28/0.07/0.35	0.41/0.51/0.38	0.29/0.40/0.02	0.37/0.34/0.34	0.38/0.37/0.45	0.24/0.33/0.54	
Median	0.40/0.50/0.61	0.29/0.38/0.62	0.26/0.26/0.55	0.39/0.48/0.67	0.35/0.37/0.32	0.37/0.35/0.44	0.40/0.32/0.55	0.27/0.39/0.59	0.33/0.39/0.42
Mean	0.41/0.48/0.61	0.32/0.41/0.59	0.25/0.24/0.52	0.39/0.43/0.61	0.35/0.37/0.29	0.39/0.34/0.43	0.39/0.32/0.52	0.31/0.36/0.59	0.32/0.36/0.39

Italic values represents stage 2.
Bold values represent stage 3.

overall κ-value for each pathologist ranged between 0.25 and 0.40 and the overall κ-value of all the pathologists was 0.34 (fair). Global κ-value was calculated to test the agreement between each pathologist's diagnosis and the majority diagnosis. It ranged from 0.39 (fair) to 0.63 (good) (Table 2).

Out of 81 cases, agreement among the majority of pathologists was observed in 34 lesions of usual ductal hyperplasia, 29 lesions of atypical ductal hyperplasia and 13 lesions of ductal carcinoma *in situ*. Equivocal agreement (cases in which an equal number of diagnoses were identified for more than one lesion) was obtained for five lesions. Table 3 shows the cumulative distribution of diagnoses reported by individual pathologists compared with the majority diagnosis for stage 1. Category specific crude agreement was 76% for usual ductal hyperplasia, 73% for ductal carcinoma *in situ* and 63% for atypical ductal hyperplasia. Overall, the percentage agreement was 70%. The category specific κ-value was lowest for atypical ductal hyperplasia (range 0.14–0.56, mean 0.43) and highest for usual ductal hyperplasia (range 0.37–0.76, mean 0.65). In stage 2 of the study, similar results were obtained as stage 1 (Tables 1, 2 and 3).

In stage 3 of the study, an ADH-5 immunostain was used along with the H&E slides of 75 cases. Complete agreement among nine pathologists was achieved in 24 (32%) cases: 23 usual ductal hyperplasia and one ductal carcinoma *in situ*. At least eight agreed in 39 (52%) cases and seven or more agreed in 47 (63%) cases. This was an improvement of agreement in 15 cases over stage 1. The majority diagnosis was 39 of usual ductal hyperplasia, 23 of atypical ductal hyperplasia and 12 of ductal carcinoma *in situ*.

The interobserver variations among the pathologists ranged from 0.02 (poor) to 0.83(very good). The mean overall κ-value for each pathologist ranged between 0.29 and 0.61 and the overall κ-value for all pathologists was 0.50 (moderate) (Table 1). There was a statistically significant improvement in the overall agreement rate between stages 1 and 3 ($P = 0.015$) (Figure 1). The global κ-values ranged from 0.42 (moderate) to 0.89 (very good) (Table 2).

Table 3 shows the cumulative distribution of diagnoses reported by individual pathologists compared with the majority diagnosis for stage 3. Category specific crude agreement was 92% for usual ductal hyperplasia, 74.1% for ductal carcinoma *in situ* and 67% for atypical ductal hyperplasia. Overall, the percentage agreement was 82%. The category specific κ-value was lowest for atypical ductal hyperplasia (range 0.21–0.84, mean 0.58) and highest for usual ductal hyperplasia (range 0.53–0.95, mean 0.81). There was a change of the majority diagnosis in seven cases from atypical ductal hyperplasia in stage 1 to usual ductal hyperplasia in stage 3 ($P = 0.0015$) (Figure 2).

The average duration between reviews of slides at any stage in this study was 4 weeks. Table 4 shows

Table 2 Category specific and global κ-values estimated between each pathologists and the majority diagnosis of breast lesions of stage 1/stage 2/stage 3

| Pathologist	UDH	ADH	DCIS	Global κ
P1	0.68/*0.92*/**0.92**	0.47/*0.82*/**0.84**	0.60/*0.84*/**0.91**	0.58/*0.86*/**0.89**
P2	0.65/*0.74*/**0.95**	0.49/*0.65*/**0.78**	0.51/*0.84*/**0.76**	0.55/*0.73*/**0.84**
P3	0.37/*0.26*/**0.84**	0.31/*0.21*/**0.64**	0.70/*0.75*/**0.63**	0.42/*0.38*/**0.72**
P4	0.76/*0.80*/**0.92**	0.52/*0.51*/**0.79**	0.54/*0.61*/**0.79**	0.62/*0.65*/**0.84**
P5	0.71/*0.73*/**0.53**	0.49/*0.58*/**0.36**	0.48/*0.78*/**0.29**	0.57/*0.69*/**0.42**
P6	0.76/*0.73*/**0.89**	0.56/*0.34*/**0.21**	0.54/*0.43*/**0.32**	0.63/*0.50*/**0.51**
P7	0.76/*0.79*/**0.81**	0.50/*0.20*/**0.21**	0.42/*0.46*/**0.70**	0.58/*0.50*/**0.70**
P8	0.65/*0.51*/**0.89**	0.14/*0.09*/**0.56**	0.35/*0.54*/**0.76**	0.39/*0.38*/**0.80**
P9	0.52/*0.62*/**0.54**	0.41/*0.44*/**0.73**	0.53/*0.70*/**0.68**	0.49/*0.57*/**0.48**
Median	0.68/*0.73*/**0.89**	0.49/*0.44*/**0.28**	0.53/*0.70*/**0.70**	0.57/*0.57*/**0.72**
Mean	0.65/*0.68*/**0.81**	0.43/*0.43*/**0.58**	0.52/*0.66*/**0.65**	0.54/*0.58*/**0.69**

Italic values represents stage 2.
Bold values represent stage 3.

Table 3 Cumulative distribution of histological diagnosis of breast lesions compared between the pathologists and the majority diagnosis of stage 1/stage 2/stage 3

All pathologists	Majority diagnosis											
	UDH (%)			ADH (%)			DCIS (%)			Total		
Stage	1	2	3	1	2	3	1	2	3	1	2	3
UDH	238 (76)	226 (76)	323 (92)	35 (14)	36 (13)	29 (14)	5 (5)	2 (1)	6 (6)	278	264	358
ADH	62 (20)	70 (24)	27 (8)	164 (63)	176 (65)	140 (68)	24 (22)	30 (20)	22 (20)	250	276	189
DCIS	15 (5)	1 (0)	1 (0)	61 (24)	58 (22)	38 (18)	79 (73)	121 (79)	80 (74)	155	180	119
Total	315	297	351	260	270	207	108	153	108	683	720	666

UDH: usual ductal hyperplasia; ADH: atypical ductal hyperplasia; DCIS: ductal carcinoma *in situ*.
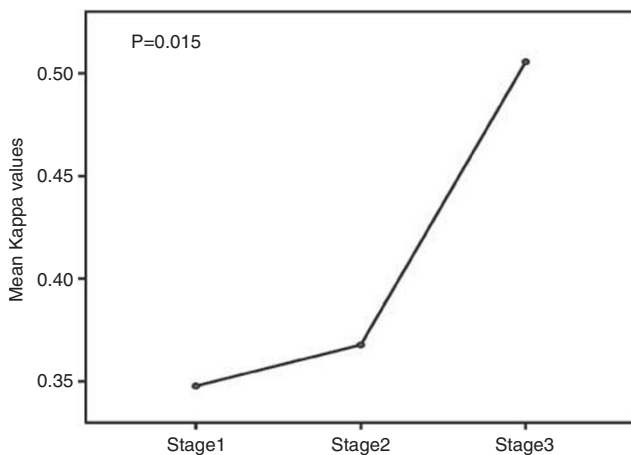
Figure 1 Graph demonstrating change in overall κ-values through each stage.

the less than perfect consistency of each pathologist in reaching the same diagnosis on rereading the same sections. The intraobserver agreement ranged from a minimum of 0.39 (fair) to a maximum of 0.88 (very good).

Discussion

This study is based on a large number of cases requiring a significant amount of pathologists' time;

all evaluations were done by the pathologists in addition to their daily busy sign-outs. The pathologists were aware that their interpretations would not have any clinical impact and it is possible that they spent significantly less time evaluating the lesions than they would in clinical practice. The artificial reading conditions, such as lack of levels and evaluation being confined to a marked area, could have affected the agreement rate in the series. The size of the lesion has been shown to be an important parameter in distinguishing atypical ductal hyperplasia from ductal carcinoma *in situ*. The use of a size criterion is strongly recommended and an integral part of the atypical ductal hyperplasia definitions proposed by Page *et al*[16] and Tavassoli *et al*.[17] In spite of these limitations, the reproducibility of breast histological diagnosis among nine pathologists was fair (κ of stages 1 and 2 = 0.34 and 0.37). These results are still within the range of observations seen in prior studies.[7–11,18–20]

The preselection of difficult/challenging cases could also have had a significant impact on the results of the study. This is illustrated by the low number of cases (11%) with complete agreement, as well as the low intraobserver agreement in stages 1 and 2 of the study. Similarly, Rosai[9] observed no agreement among the cases seen by five experienced pathologists in his study. On the other hand, Wells *et al*[11] used a representative sample of diagnostic categories seen routinely in general practice and
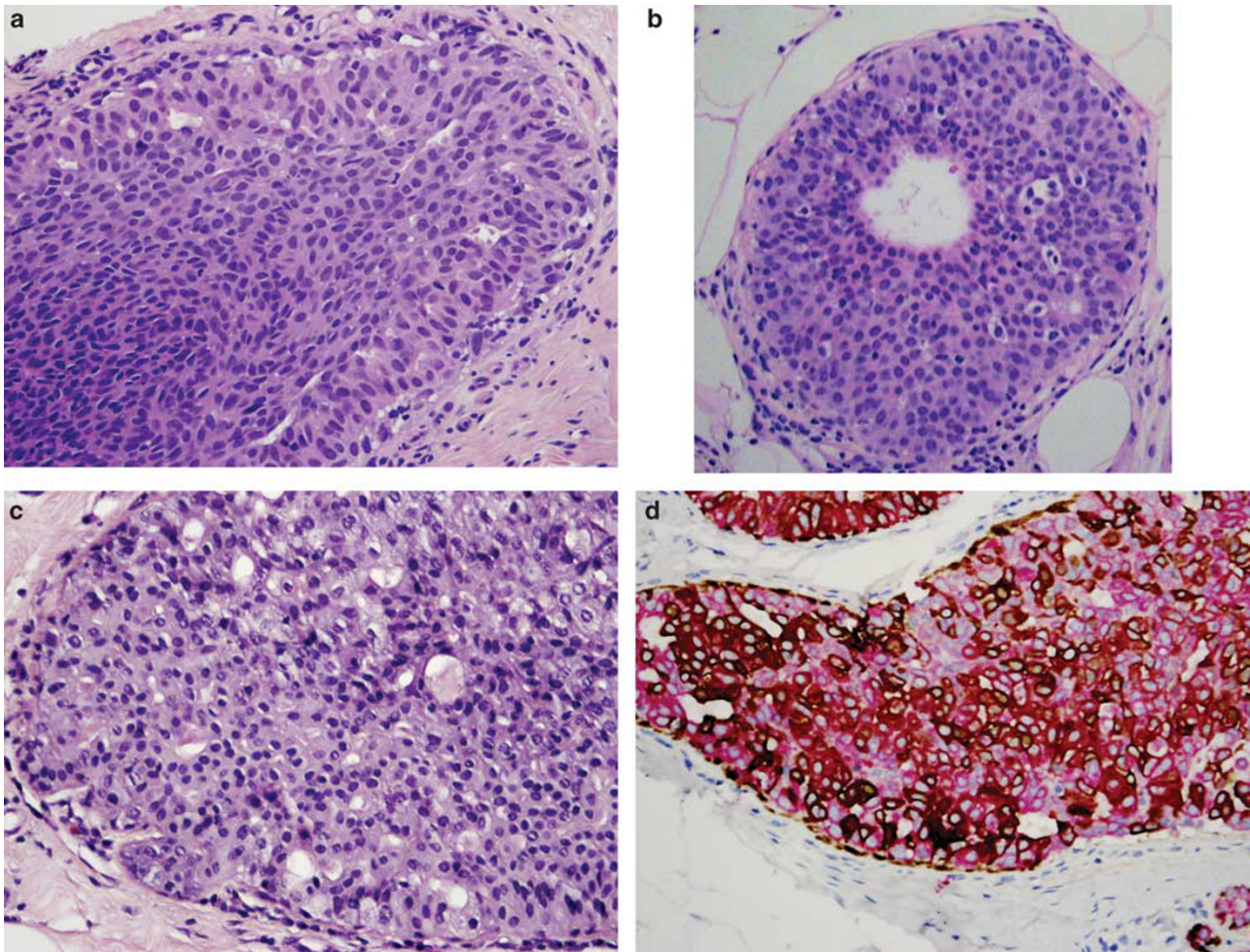
Pathologists' concordance in diagnosis of proliferative breast lesions
RK Jain *et al*
921

Figure 2 In (**a**), four pathologists interpreted this slide as usual ductal hyperplasia, three as atypical ductal hyperplasia and two as ductal carcinoma *in situ*. In (**b**), two pathologists called it usual ductal hyperplasia, three as atypical ductal hyperplasia and four as ductal carcinoma *in situ*. (**c**, **d**) represent the same lesion, where in stage 1, seven pathologists called it atypical ductal hyperplasia, while after using immunohistochemistry, all nine pathologists called it usual ductal hyperplasia.

Table 4 Intraobserver variation among nine pathologists in three stages (κ-value)

Pathologists	Stage 1 vs 2	Stage 1 vs 3	Stage 2 vs 3
P1	0.88	0.66	0.72
P2	0.66	0.53	0.68
P3	0.39	0.30	0.16
P4	0.59	0.48	0.65
P5	0.63	0.28	0.34
P6	0.66	0.47	0.39
P7	0.58	0.60	0.55
P8	0.61	0.48	0.34
P9	0.56	0.42	0.52

observed a higher level of agreement ($\kappa = 0.71$) among the participating community pathologists.

In our analyses of the intraobserver variability, the agreement rate (κ) ranged from 0.39 to 0.88. Beck[7] observed an overall agreement of 78% (κ not provided) in individual diagnoses of the pathologists. Most of the inconsistencies in the current

study and their study were due to borderline lesions. It is unlikely that with a large number of cases and the relative long duration between reads (average, 4 weeks) that 'memory' would have contributed to the intraobserver reproducibility.

Schnitt *et al*[10] concluded that the interobserver variability could be reduced with the use of standardized criteria for ductal lesions. By using Page's criteria and providing training slides, they observed 58% complete agreement among the participating pathologists. In contrast, Palazzo and Hyslop[20] documented a low κ-value (0.36) in the diagnosis of benign and malignant ductal lesions when their study participants (community and academic pathologists) used the same standardized criteria. In our study, the participants were asked to use their own criteria, which they use in their daily practice, and no teaching slides were provided.

In the current study, a moderate level of agreement ($\kappa = 0.54$) was achieved for all diagnostic categories. Among seven of nine observers, there was a relatively good agreement in the diagnosis of

usual ductal hyperplasia. However, in the two intermediate categories (atypical ductal hyperplasia and ductal carcinoma *in situ*), there were disagreements resulting in κ-values between fair to moderate. Most of the studies investigating concordance rates documented that high interobserver variation was mostly due to problems in differentiating atypical ductal hyperplasia and low-grade ductal carcinoma *in situ*.[8,11,19–22] The category specific κ-value was lowest for atypical ductal hyperplasia (0.43 for stages 1 and 2) in this study. These results are similar to studies by Palli et al[8] and MacGrogan et al,[23] with the lowest category specific κ-values for the diagnosis of atypical ductal hyperplasia (0.38 and 0.36, respectively). We agree with Elston et al,[21] who stated that the poor consistencies observed in the diagnosis of atypical ductal hyperplasia lesions raises serious concerns regarding the robustness of the current diagnostic criteria. Their use of digitized images serving the function of marked specific fields did not improve the κ-values.

In order to improve the concordance rate, we used a recently commercialized immunohistochemical breast marker cocktail (combination of CK5, 14, 7, 18 and p63) antibody. Myoepithelial/basal cells express CK5, 7, 14, 17 and other specific markers such as smooth muscle actin, calponin and p63, while luminal cells express keratins such as 7, 8, 18 and 19.[24–28] In usual ductal hyperplasia, with variable architectural and cellular features, cytokeratins, particularly basal types, are stained heterogeneously showing a mosaic pattern. In contrast, low-nuclear-grade ductal carcinoma *in situ* stains positively for CK8/18 and CK19, while it is negative for CK5/6 and/or CK14. These features are highlighted by the use of high-molecular weight cytokeratins like 34βE12, CK5/6 and CK14.[25,29,30] With the use of this combination of high- and low-molecular weight cytokeratin antibodies along with the H&E slides, we observed significant improvement in the concordance rate among pathologists from fair (0.34 of stage 1) to moderate (0.50 of stage 3). Similar to our study, Douglas-Jones et al[31] have reported an improvement in the diagnostic agreement of core biopsy specimens with the use of immunohistochemistry for CK5/6, calponin and p63. In contrast, MacGrogan et al[23] have not been able to show significant improvement in the concordance rate ($\kappa = 0.58$) by CK5/6 and E-cadherin.

Apart from improving the concordance rate, we also observed a significant reduction in the number of atypical ductal hyperplasia diagnoses with the immunostain. Prior studies have demonstrated that 40% of lesions diagnosed as atypical ductal hyperplasia on core biopsies consisted only of epithelial hyperplasia or other benign lesions without atypia on excision.[32] This highlights an important issue of overdiagnosis and misclassification of atypical ductal hyperplasia, which has a different treatment protocol compared with benign lesions. Misclassification of benign lesions as atypical or malignant

results in excessive patient anxiety and treatment costs. On the other hand, misdiagnosing a malignant tumor as a benign tumor leads to inadequate treatment. This misclassification was clearly demonstrated in a large screening study (NCI—American Cancer Society) where 9% of women who were being treated for noninfiltrating carcinoma did not have a malignant lesion.[33–36] Our study achieved a substantial decrease (8%) in the number of atypical ductal hyperplasia diagnoses after the use of immunostains. These lesions were equivocally diagnosed in the benign category in stage 3.

Several criteria to differentiate these lesions exist; however, it is not clear as to which criteria to apply and what is the relative 'weightage' given to the different features. Optimal tissue fixation and processing has also been identified as major factors in reducing interobserver variation in the histologic grading of breast carcinomas.[11] Formation of a consensus building committee or review of all the pathologic material through a central laboratory or headquarters could also improve the concordance rate,[37] but is not practical. External quality assessment, rereading or second evaluation of the slides, examining further material including deeper levels and additional tissue blocks, where appropriate, could also improve the consistency. Immuohistochemical stains like the breast cocktail marker in the current study could also help in improving the agreement rate and reduce overdiagnosis of atypical ductal hyperplasia lesions. Newer technologies like computer-aided diagnosis after validation could assist pathologists in the analysis of the slides and improve the diagnosis and management of intraductal breast lesions.[38]

In summary, we have shown that the diagnostic agreement for noninvasive epithelial breast proliferations based on morphology is fair and it significantly improved by the combination of high- and low-molecular weight cytokeratins immunostain.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Ernster VL, Barclay J. Increases in ductal carcinoma *in situ* (DCIS) of the breast in relation to mammography: a dilemma. J Natl Cancer Inst Monogr 1997;(22):151–156.

2 Ashikari R, Huvos AG, Snyder RE, et al. Proceedings: a clinicopathologic study of atypical lesions of the breast. Cancer 1974;33:310–317.

3 Ashikari R, Huvos AG, Snyder RE, et al. A clinicopathologic study of atypical lesions of the breast further follow up. Pathol Res Pract 1980;166:481–490.

4 Goldenberg VE, Goldenberg NS, Sommers SC. Comparative ultrastructure of atypical ductal hyperplasia, intraductal carcinoma, and infiltrating ductal carcinoma of the breast. Cancer 1969;24:1152–1169.

5 Wellings SR, Jensen HM, Marcum RG. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. J Natl Cancer Inst 1975;55:231–273.

6 Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. N Engl J Med 1985;312:146–151.

7 Beck JS. Observer variability in reporting of breast lesions. J Clin Pathol 1985;38:1358–1365.

8 Palli D, Galli M, Bianchi S, *et al.* Reproducibility of histological diagnosis of breast lesions: results of a panel in Italy. Eur J Cancer 1996;32A:603–607.

9 Rosai J. Borderline epithelial lesions of the breast. Am J Surg Pathol 1991;15:209–221.

10 Schnitt SJ, Connolly JL, Tavassoli FA, *et al.* Inter-observer reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am J Surg Pathol 1992;16:1133–1143.

11 Wells WA, Carney PA, Eliassen MS, *et al.* Statewide study of diagnostic agreement in breast pathology. J Natl Cancer Inst 1998;90:142–145.

12 Simpson JF. Update on atypical epithelial hyperplasia and ductal carcinoma *in situ*. Pathology 2009;41: 36–39.

13 Pinder SE, Ellis IO. The diagnosis and management of pre-invasive breast disease: ductal carcinoma *in situ* (DCIS) and atypical ductal hyperplasia (ADH)—current definitions and classification. Breast Cancer Res 2003;5:254–257.

14 Fleiss JL. Statistical Methods for Rates and Proportions Vol., John Wiley and Sons: New York, 1981.

15 Carpenter CR. Kappa statistic. CMAJ 2005;173: 15–16; author reply 7.

16 Page DL, Dupont WD, Rogers LW, *et al.* Atypical hyperplastic lesions of the female breast. A long-term follow-up study. Cancer 1985;55:2698–2708.

17 Tavassoli FA, Norris HJ. A comparison of the results of long-term follow-up for atypical intraductal hyperplasia and intraductal hyperplasia of the breast. Cancer 1990;65:518–529.

18 Bodian CA, Perzin KH, Lattes R, *et al.* Reproducibility and validity of pathologic classifications of benign breast disease and implications for clinical applications. Cancer 1993;71:3908–3913.

19 Ghofrani M, Tapia B, Tavassoli FA. Discrepancies in the diagnosis of intraductal proliferative lesions of the breast and its management implications: results of a multinational survey. Virchows Arch 2006;449: 609–616.

20 Palazzo JP, Hyslop T. Hyperplastic ductal and lobular lesions and carcinomas *in situ* of the breast: reproducibility of current diagnostic criteria among community- and academic- based pathologists. Breast J 1998; 4:230–237.

21 Elston CW, Sloane JP, Amendoeira I, *et al.* Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast. European commission working group on breast screening pathology. Eur J Cancer 2000;36:1769–1772.

22 Bianchi S, Palli D, Galli M, *et al.* Reproducibility of histological diagnoses and diagnostic accuracy of non palpable breast lesions. Pathol Res Pract 1994; 190:69–76.

23 MacGrogan G, Arnould L, de Mascarel I, *et al.* Impact of immunohistochemical markers, CK5/6 and E-cadherin on diagnostic agreement in non-invasive proliferative breast lesions. Histopathology 2008;52: 689–697.

24 Yeh IT, Mies C. Application of immunohistochemistry to breast lesions. Arch Pathol Lab Med 2008;132: 349–358.

25 Otterbach F, Bankfalvi A, Bergner S, *et al.* Cytokeratin 5/6 immunohistochemistry assists the differential diagnosis of atypical proliferations of the breast. Histopathology 2000;37:232–240.

26 Heatley M, Maxwell P, Whiteside C, *et al.* Cytokeratin intermediate filament expression in benign and malignant breast disease. J Clin Pathol 1995;48:26–32.

27 Purkis PE, Steel JB, Mackenzie IC, *et al.* Antibody markers of basal cells in complex epithelia. J Cell Sci 1990;97(Part 1):39–50.

28 Wetzels RH, Holland R, van Haelst UJ, *et al.* Detection of basement membrane components and basal cell keratin 14 in noninvasive and invasive carcinomas of the breast. Am J Pathol 1989;134:571–579.

29 Lacroix-Triki M, Mery E, Voigt JJ, *et al.* Value of cytokeratin 5/6 immunostaining using D5/16 B4 antibody in the spectrum of proliferative intraepithelial lesions of the breast. A comparative study with 34betaE12 antibody. Virchows Arch 2003;442: 548–554.

30 Moinfar F, Man YG, Lininger RA, *et al.* Use of keratin 35betaE12 as an adjunct in the diagnosis of mammary intraepithelial neoplasia-ductal type–benign and malignant intraductal proliferations. Am J Surg Pathol 1999;23:1048–1058.

31 Douglas-Jones A, Shah V, Morgan J, *et al.* Observer variability in the histopathological reporting of core biopsies of papillary breast lesions is reduced by the use of immunohistochemistry for CK5/6, calponin and p63. Histopathology 2005;47:202–208.

32 Gal-Gombos EC, Esserman LE, Recine MA, *et al.* Large-needle core biopsy in atypical intraductal epithelial hyperplasia including immunohistochemical expression of high molecular weight cytokeratin: analysis of results of a single institution. Breast J 2002; 8:269–274.

33 Beahrs OH, Shapiro S, Smart C. Report of the working group to review the National Cancer Institute—American Cancer Society breast cancer detection demonstartion projects. J Natl Cancer Inst 1979;62: 640–709.

34 Beahrs OH, Smart CR. Diagnosis of minimal breast cancers in the BCDDP: the 66 questionable cases. Cancer 1979;43:848–850.

35 Beahrs OH. Early detection of breast cancer: results of a screening programme. Ann R Coll Surg Engl 1980;62: 38–40.

36 Smart CR, Byrne C, Smith RA, *et al.* Twenty-year follow-up of the breast cancers diagnosed during the Breast Cancer Detection Demonstration Project. CA Cancer J Clin 1997;47:134–149.

37 Fisher ER, Costantino J. Quality assurance of pathology in clinical trials. The National Surgical Adjuvant Breast and Bowel Project experience. Cancer 1994;74: 2638–2641.

38 Dundar MM, Badve S, Jain R, *et al.* Computerized analysis of breast microscopic tissues for improved classification of intraductal lesions. IEEE Trans Biomed Eng; 4 February 2011 [e-pub ahead of print].