J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Research Article

Quantitative nucleic features are effective for discrimination of intraductal proliferative lesions of the breast

Masatoshi Yamada[1], Akira Saito[2], Yoichiro Yamamoto[3], Eric Cosatto[4], Atsushi Kurata[1], Toshitaka Nagao[5], Ayako Tateishi[6], Masahiko Kuroda[1]

Departments of [1]Molecular Pathology, [2]Quantitative Pathology and Immunology and [5]Human Pathology, Tokyo Medical University, Tokyo, [3]Department of Pathology, School of Medicine, Shinshu University, [6]Department of Laboratory Medicine, Shinshu University Hospital, Nagano, Japan, [4]Department of Machine Learning, NEC Laboratories America, Princeton, NJ, USA

E-mail: *Dr. Akira Saito - ak-saito@tokyo-med.ac.jp, *Prof. Masahiko Kuroda - kuroda@tokyo-med.ac.jp
*Corresponding author

Masatoshi Yamada and Akira Saito are equly contributed to this work.

Received: 19 June 2015 Accepted: 11 Novemebr 2015 Published: 29 January 2016

Abstract

Background: Intraductal proliferative lesions (IDPLs) of the breast are recognized as a risk factor for subsequent invasive carcinoma development. Although opportunities for IDPL diagnosis have increased, these lesions are difficult to diagnose correctly, especially atypical ductal hyperplasia (ADH) and low-grade ductal carcinoma *in situ* (LG-DCIS). In order to define the difference between these lesions, many molecular pathological approaches have been performed. However, still we do not have a molecular marker and objective histological index about IDPLs of the breast. **Methods:** We generated full digital pathology archives from 175 female IDPL patients, including usual ductal hyperplasia (UDH), ADH, LG-DCIS, intermediate-grade (IM)-DCIS, and high-grade (HG)-DCIS. After total 2,035,807 nucleic segmentations were extracted, we evaluated nuclear features using step-wise linear discriminant analysis (LDA) and a support vector machine. **Results:** High diagnostic accuracy (81.8–99.3%) was achieved between pathologists' diagnoses and two-group LDA predictions from nucleic features for IDPL discrimination. Grouping of nuclear features as size and shape-related or intranuclear texture-related revealed that the latter group was more important when distinguishing between normal duct, UDH, ADH, and LG-DCIS. However, these two groups were equally important when discriminating between LG-DCIS and HG-DCIS. The Mahalanobis distances between each group showed that the smallest distance values occurred between LG-DCIS and IM-DCIS and between ADH and Normal. On the other hand, the distance value between ADH and LG-DCIS was larger than this distance. **Conclusions:** In this study, we have presented a practical and useful digital pathological method that incorporates nuclear morphological and textural features for IDPL prediction. We expect that this novel algorithm is used for the automated diagnosis assisting system for breast cancer.

Key words: Intraductal proliferative lesion of breast, nucleic analysis, whole slide imaging

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.175380

Quick Response Code:

This article may be cited as: Yamada M, Saito A, Yamamoto Y, Cosatto E, Kurata A, Nagao T, *et al*. Quantitative nucleic features are effective for discrimination of intraductal proliferative lesions of the breast. J Pathol Inform 2016;7:1.
Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2016/7/1/1/175380

INTRODUCTION

Before the most recent quarter-century, the majority of breast cancer cases were detected at an invasive stage. The

dissemination of simple mammography-based screening methods has dramatically increased the opportunity for diagnosing ductal carcinoma *in situ* (DCIS). The development of mammography has led to improved detection of intraductal proliferative lesions (IDPLs), including usual ductal hyperplasia (UDH), atypical ductal hyperplasia (ADH), and DCIS, which is considered the precursor lesion of invasive cancer.

Furthermore, great progress has been achieved in breast cancer management. Previously, radical mastectomy was the main treatment for breast cancer, but now the use of limited surgery in combination with radiation and chemotherapy has increased. Earlier breast lesion detection has lowered the patient age and accordingly increased expectations for limited surgery from an esthetic viewpoint. Given these parameters, appropriate pathological diagnoses of breast lesions, as well as precursor lesions are now in greater demand than ever before.

Although the definitive diagnosis of a breast lesion depends on a histopathological diagnosis via biopsy, the rate of diagnostic agreement among pathologists is only moderate for IDPLs, especially ADH and low-grade (LG)-DCIS.[1-4] Therefore, the establishment of objective diagnostic criteria is an urgent requirement because the number of ADH diagnoses in clinical practice increases annually.[5] Initially, ADH was described as ductal hyperplasia with "a loss of shape" and was not thought to be associated with the transition to invasive cancer.[6] ADH was subsequently found to confer a 4–5-fold greater risk of transition to invasive cancer relative to UDH, and it has become necessary to consider ADH and UDH as independent lesions.[7,8] The recent but widely accepted histological definition of ADH is a lesion that has both cytological and architectural features of LG, noncomedo DCIS but involves only a single ductal space or is <2 mm in size.[8,9] The distinction between ADH and LG-DCIS, however, remains difficult and diagnostic dissociation among pathologists is not uncommon.

The pathological concept of "ductal intraepithelial neoplasia (DIN)" was proposed in 1997 as a solution to this problem.[10] DIN explains the progression of IDPLs from UDH to DCIS as a sequential lesion. DIN can be classified into three categories. Grade 1 DIN (DIN1) includes UDH, ADH, and LG-DCIS. Grade 2 (DIN2), and Grade 3 DIN (DIN3) correspond to intermediate-grade (IM)-DCIS and high-grade (HG)-DCIS, respectively. Hence, the most significant characteristic of the DIN classification is the inclusion of ADH and LG-DCIS in the same category, given their low differential diagnostic agreement. The risks of invasive cancer development from ADH and DCIS are relatively low (4–5% and 8–10%, respectively),[11] and the psychological strain upon patients and their families can be reduced by avoiding the use of the word "carcinoma" for cases of DCIS.

On the other hand, a clinical gap remains between ADH and DCIS. The topics of treatment following ADH diagnosis by needle biopsy and the implication of ADH in the resection stump of a breast surgery specimen remain open for discussion; therefore, the need for accurate ADH and DCIS diagnosis is considered significant. With the recent advances in molecular biology, gene expression profiling has been used to explain carcinogenic mechanisms and facilitate diagnoses. For breast cancer, prognosis prediction tools such as OncotypeDX® (Genomic Health, Inc., Redwood City, CA, USA) have been commercialized. However, given their high cost, it is unlikely that such tools will be widely used. In addition, although genetic analysis derives information from a very small number of cells in a cancer tissue, pathologists make diagnoses based on a comprehensive analysis of all tissues in a section. From their perspective, a more comprehensive range of quantitative information should be used to facilitate diagnosis. Hence, according to pathologists' experiences,[12] whole slide imaging (WSI)-based analysis is the best tool for objectively evaluating the morphology of lesions in pathological tissue specimens. In addition to the molecular pathological approach, this technique considers the integration of a series of information related to molecular expression, nuclear morphology, and the organizational structure of the pathological malignancy. In this study, the usefulness of image analysis is verified using WSI of pathological sections of mammary gland ductal lesions.

METHODS

Samples

We analyzed a total of 175 breast tissue specimens representative of 6 histological types [Table 1]: 4 UDH specimens [Figure 1a], 21 ADH specimens [Figure 1b],

Table 1: Number of measured nuclei according to histological diagnosis

	Case number	ROIs	Measured nuclei
Normal	25	170	189,843
UDH	4	16	23,808
ADH	21	140	126,708
LG-DCIS	72	491	929,803
IM-DCIS	29	351	595,821
HG-DCIS	24	251	169,824
Total	175	1419	2,035,807

ROIs: Regions of interest, UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

72 LG-DCIS specimens [Figure 1c], 29 IM-DCIS specimens, and 24 HG-DCIS specimens [Figure 1d], as well as 25 normal breast tissue specimens obtained from formalin-fixed, paraffin-embedded (FFPE) blocks in which no diagnostic lesions had been detected. All specimens were diagnosed and surgically obtained at Shinshu University Hospital between 2011 and 2013. None of the specimens included invasive lesions nor those treated with neoadjuvant chemotherapy. This study was performed according to the Helsinki Declaration and was approved by the ethical committee of Shinshu University, Japan.

Tissue Preparation and Whole Slide Scanning

All FFPE samples were sectioned at a thickness of 3 μm. After hematoxylin and eosin staining according to the standard method, all slides were scanned using a WSI scanner (Nanozoomer 2.0-HT slide scanner; Hamamatsu Corp., Hamamatsu, Shizuoka, Japan) at ×20 magnification and were stored as TIFF files on a computer system.

Analytical Image Selection and Histological Classification

From whole slide images, analysis target areas or regions of interest (ROI) were selected manually. A single ROI image measured 2048 pixels × 2048 pixels, corresponding to 0.942 mm × 0.942 mm on the slide. Whole slide images contained mammary ducts, as well as stromal cells and areas of lymphocyte aggregation, among other features. Manual ROI selection was performed to confirm the positions of mammary glands and avoid areas containing scanning artifacts (e.g., poor focus). A total of 1419 ROIs were selected [Table 1]. For each ROI image, we diagnosed lesions based on the World Health Organization (WHO) classification criteria. At least three well-trained pathologists independently diagnosed and scored each ROI; collective consensus regarding discrepant results was reached through discussion. We initially classified each ROI into 1 of 4 main classes: Normal, UDH, ADH, and DCIS. DCIS ROIs were subsequently classified as LG, IM-grade, and HG.

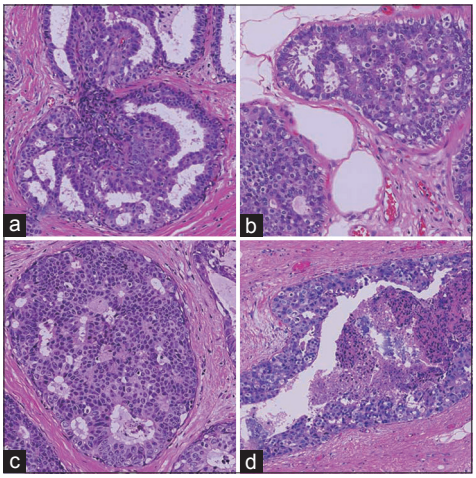

Figure 1: Microscopic morphology of hematoxylin and eosin stained intraductal proliferative lesions (×200). (a) Usual ductal hyperplasia, (b) atypical ductal hyperplasia, (c) low-grade ductal carcinoma *in situ*, and (d) high-grade ductal carcinoma *in situ*

Quantitative Morphological Image Analysis

We subjected the ROI images to a quantitative morphological analysis according to a previously described method.[13] The first analytical step was the extraction of the nuclear contours. Three image filters (2 difference-of-Gaussian sizes; 1 Hough) were used to locate the approximate centers of nuclei from hematoxylin signal images. After aggregation and nonmaxima suppression, polar cross-sections were extracted from the candidate centers. From the hematoxylin signal peaks on polar images, an algorithm determined the nuclear contour line using a snake line adaptation method. Figure 2 shows an example of nuclear extraction. Although the ROIs were centered on areas of mammary gland lesions, the images still contained fibroblasts, myoepithelial cells, lymphocytes, and other immunological cells. These cells were excluded using manually created masks. The algorithm then measured textural features using the integrated diffusion gradient method. For each nucleus, the algorithm measured a total of 40 morphological features [Table 2]. These features were separated into two Groups: Size- and shape-related features and intranuclear texture features.

For each ROI, the analysis could produce 200–5000 nuclei; from all ROIs, a total of 2,035,808 nuclei were measured. The algorithm then summarized these nuclear features by evaluating their statistical distributions over each ROI. The average, variance, standard deviation, median, mode, and percentiles

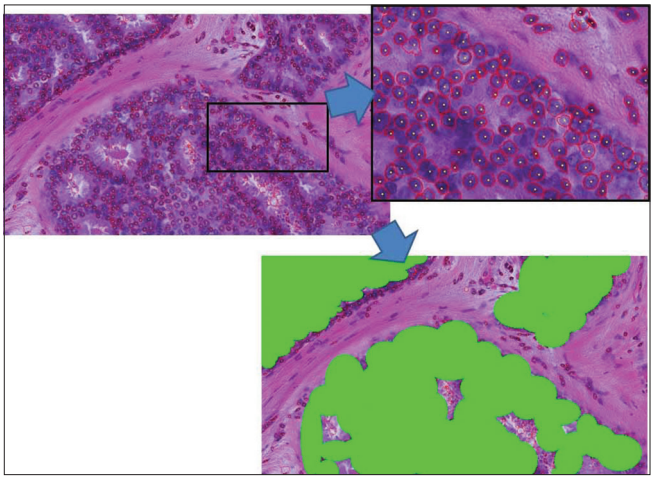

Figure 2: Example of nuclear contour extraction results. The enlarged partial position is on the upper right. Red lines indicate the automatically extracted nuclear contour line. Yellow dots indicate the nuclear center position. The lower image is a manually created masked image. Nuclear features were measured only on selected nuclei indicated in green areas

(10, 30, 50, 70, and 90) were measured. To reduce the effects of nuclear contour extraction errors (e.g., ≥2 touching nuclei resulting in a single contour), the algorithm also measured these statistical distributions over 80% of the centered nuclei by ignoring the 10% comprising the largest nuclei and 10% comprising the smallest nuclei. A total of 15 statistics were measured for each of 40 features, resulting in 600 ROI features. A first feature reduction was then applied by eliminating features that exhibited no variance over the 1419 ROI dataset, yielding 472 features.

To identify the feature sets that were useful for distinguishing each lesion, a multiclass step-wise linear discriminant analysis (LDA) method was used. For this feature selection, a *P*- 0.05 was set for the discriminant function IN and OUT level. We used the Statistical Analysis Software Package R (R Project for Statistical Computing; http://www.r-project.org).

Furthermore, to confirm the discrimination level, we applied the machine learning method support vector machine (SVM), for which we used both linear and radial basis function kernels.[14] In this paper, we have reported the results using linear kernel SVM. To achieve a more accurate discrimination, we trained the SVM with 98% of the data and tested it with the remaining 2% (split randomly). This was performed >100 times, and the average results are reported. Although very similar to LDA, the advantage of linear SVM is that it looks directly at data points instead of approximate normal distributions when building the decision boundaries. Both LDA and SVM are used for discrimination purpose, but SVM needs the training set for creating the model, and prediction set for checking. SVM uses all given features, on the other hand, step-wise LDA uses minimum features set. Comparing both method results and discrimination rate, we analyzed features contribution level for each lesion discrimination.

Table 2: Nuclear morphological parameters

Nucleus size and shape parameters			
Nucleus area size	Nucleus contour line length	Roundness	Long axis length
Short axis length	Ellipsoidal ratio	Contour line complexity	IDG1 area size
Nucleus density in tissue	Nuclear arrayment level		
Intranuclear texture parameters			
GLCM angular 2nd moment	GLCM contrast	GLCM homogeneity	GLCM entropy
IDG2 nucleus volume	IDG3-8	IDG9-14	IDG15-20
IDG7	IDG10	IDG11	IDG12
IDG13	IDG14	IDG15	IDG16
IDG21-26	Nuclear texture complex		

Average; variance; standard deviation; median, mode; 10%, 30%, 50% 70%, and 90% tile data; 80% based average variance, standard deviation, median, mode. GLCM: Gray level co-occurrence matrix, IDG: Integrated diffusion gradient, IDG1: Ratio of nucleus area size and rectangle box area (long × short axis), IDG2: Ratio of nucleus 3D volume to cuboid volume, IDG3-8: Total volume over 6 threshold intensity levels, IDG9-14: Increased volume over each threshold intensity level, IDG15-20: Counts for each threshold intensity level cluster, IDG 21-26: Image fractal dimensions for each threshold intensity level

RESULTS

We first applied LDA and SVM to all 6 histological groups: Normal, UDH, ADH, LG-DCIS, IM-DCIS, and HG-DCIS [Tables 3a and b]. Table 3a shows an analysis of a total of 170 Normal mammary gland ROIs. By step-wise LDA, 147 ROIs were correctly classified as Normal. The remaining 4, 10, 5, 3, and 1 ROIs were misclassified, respectively, as UDH, ADH, LG-DCIS, IM-DCIS, and HG-DCIS. The diagonal cells in the table show the numbers of correctly classified ROIs. The overall accuracy rates of LDA and SVM were 77.4–95.9%, respectively. LDA yielded a lower accuracy because this method assumes a normal data distribution; however, realistically the distribution is closer to a Chi-square distribution with 2° of freedom. When using LDA, ADH had the lowest accuracy level (67.1%) [Table 3a], whereas when using SVM, IM-DCIS had the lowest accuracy level (93.4%) [Table 3b].

Table 3a: Step-wise linear discriminant analysis results

	Prediction						
	Normal (%)	**UDH (%)**	**ADH (%)**	**LG-DCIS (%)**	**IM-DCIS (%)**	**HG-DCIS (%)**	**Total**
Truth							
Normal	147 (86.5)	4 (2.4)	10 (5.9)	5 (2.9)	3 (1.8)	1 (0.6)	170
UDH	0 (0)	15 (93.8)	0 (0)	1 (6.3)	0 (0)	0 (0)	16
ADH	21 (15)	5 (3.6)	94 (67.1)	11 (7.9)	8 (5.7)	1 (0.7)	140
LG-DCIS	16 (3.3)	13 (2.6)	32 (6.5)	364 (74.1)	63 (12.8)	3 (0.6)	491
IM-DCIS	8 (2.3)	3 (0.9)	8 (2.3)	69 (19.7)	257 (73.2)	6 (1.7)	351
HG-DCIS	2 (0.8)	0 (0)	3 (1.2)	14 (5.6)	11 (4.4)	221 (88.0)	251
Total	194	40	147	464	342	232	1419

Total accuracy=77.4%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

Table 3b: Linear SVM analysis results

	Prediction						
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth							
Normal	168 (98.9)	0 (0)	0 (0)	2 (1.0)	0 (0)	(0)	170
UDH	0 (0)	16 (100)	0 (0)	0 (0)	0 (0)	0 (0)	16
ADH	0 (0)	0 (0)	135 (96.4)	4 (2.9)	1 (0.7)	1 (0.7)	140
LG-DCIS	3 (0.6)	1 (0.2)	9 (1.8)	464 (94.1)	11 (2.2)	3 (0.6)	491
IM-DCIS	0 (0)	0 (0)	1 (0.2)	22 (6.2)	328 (93.4)	0 (0)	351
HG-DCIS	0 (0)	0 (0)	0 (0)	1 (0.4)	0 (0)	250 (99.6)	251
Total	171	17	145	493	340	253	1419

Total accuracy=95.9%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*, SVM: Support vector machine

In Supplementary Table 1, we report the standardized coefficient values of selected features for each pair-wise discriminant function. These values are useful for intuitively confirming the discrimination power of individual features. However, because the discriminant functions are pair-wise, for our six classes 15 coefficients required evaluation and the number of selected features became large. Hence, we performed a step-wise LDA (in which features with the highest prediction power were selected greedily and their variances were removed until the correlations stopped changing significantly) for each combination of histopathological conditions. High diagnostic accuracy (81.8–99.3%) was achieved, and the results are shown in Tables 4a and b. Table 4c shows the Mahalanobis distances between each group centroid. In this table, we note that the smallest distance values occurred between LG-DCIS and IM-DCIS and between ADH and Normal (1.73567 and 1.98269, respectively). The distance value between ADH and LG-DCIS (2.48288) was larger than these distance, and the values between HG-DCIS and the other types were, relatively, large. In Supplementary Table 2, we report the standardized coefficient values of selected features after performing step-wise LDA. Blank table cells correspond to features that were not selected by a paired histological criteria analysis.

While focusing on LG-DCIS and benign lesions, we reduced the number of groups to 4 (Normal, UDH, ADH, and LG-DCIS). Tables 5a and b present the results obtained when using all features with LDA and SVM, respectively. Tables 5c and d show the results obtained using only nucleus size and shape-related features, whereas Tables 5e and f show those based on only intranuclear textural features. We noted that using only nucleus size- and shape-related features resulted in lower accuracy rates for both LDA (62.1% vs. 81.4%, respectively) and SVM (80.0% vs. 99.8%).

The same analysis was performed while focusing on the 3 DCIS grades (LG-DCIS, IM-DCIS, and HG-DCIS), and these results are reported in Table 6a-f. For this grouping, no significant differences in accuracy were observed between the two types of features. This demonstrates that although intranuclear texture features were more important when analyzing LG-DCIS and benign lesions, the nucleus size and shape features were equally important when grading DCIS lesions.

Table 7 shows the levels of contribution for the top 22 features in an SVM analysis of the six histological classes. Nucleus size- and shape-related features accounted for 10.5%, whereas intranuclear texture accounted for 89.5%.

DISCUSSION

In the WHO classification published in 2012, IDPLs were regarded as an independent category that included UDH, columnar cell lesions (e.g., columnar cell changes and hyperplasia, flat epithelial hyperplasia), ADH, and DCIS.[15] This categorization stratified IDPLs, which carry a risk of invasive ductal carcinoma, while targeting a practical diagnosis. ADH was first proposed as a lesion harboring the risk of invasive ductal carcinoma by Page *et al*. in 1985.[7,16]

On the other hand, a relatively low rate of concordance between pathologists has been noted with respect to ADH diagnostic criteria.[1-3] Some authors have reported that the introduction of unified diagnostic criteria[17] and supportive immunohistochemistry[3] have led to an increase in diagnostic concordance. However, other authors reported that the incidence of diagnostic discrepancy was not reduced, despite the introduction of unified diagnostic criteria.[18]

Under these circumstances, Rosai proposed the addition of mammary intraepithelial neoplasia, which corresponds to all IDPLs.[19] This new classification is effective for pathologists; however, it could cause confusion with regard to clinical management. Despite its noninvasive nature, the management of DCIS, generally, conforms to that of invasive carcinoma; therefore, total tumor resection is usually applicable. Accordingly, the introduction of

Table 4a: Step-wise discriminant analyses for each combination of histopathological conditions

	Prediction		
	Normal (%)	**UDH (%)**	**Total**
Truth			
Normal	165 (97.1)	5 (2.1)	170
UDH	0 (0)	16 (100)	16
Total	165	21	186

Accuracy=97.3%, Mahalanobis' distance=3.87675. UDH: Usual ductal hyperplasia

	Prediction		
	Normal (%)	**UDH (%)**	**Total**
Truth			
Normal	149 (87.6)	21 (12.4)	170
ADH	26 (18.6)	114 (81.4)	140
Total	175	135	310

Accuracy=84.8%, Mahalanobis' distance=1.98269. UDH: Usual ductal hyperplasia

	Prediction		
	Normal (%)	**UDH (%)**	**Total**
Truth			
Normal	152 (89.4)	18 (10.6)	170
LG-DCIS	28 (5.7)	463 (94.3)	491
Total	180	481	661

Accuracy=93.0%, Mahalanobis' distance=3.25961. UDH: Usual ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

	Prediction		
	Normal (%)	**IM-DCIS (%)**	**Total**
Truth			
Normal	167 (98.2)	3 (1.8)	170
IM-DCIS	8 (2.3)	343 (97.7)	351
Total	175	21	521

Accuracy=97.9%, Mahalanobis' distance=4.12958. IM-DCIS: Intermediate-grade ductal carcinoma *in situ*

	Prediction		
	Normal (%)	**HG-DCIS (%)**	**Total**
Truth			
Normal	165 (97.1)	5 (2.9)	170
HG-DCIS	5 (2.0)	246 (98.0)	251
Total	170	251	421

Accuracy=97.6%, Mahalanobis' distance=4.36597. HG-DCIS: High-grade ductal carcinoma *in situ*

	Prediction		
	UDH (%)	**ADH (%)**	**Total**
Truth			
UDH	15 (93.8)	1 (6.2)	16
ADH	19 (13.6)	121 (86.4)	140
Total	34	122	156

Accuracy=87.2%, Mahalanobis' distance=2.24234. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia

Contd...

Table 4a: Contd...

	Prediction		
	UDH (%)	**LG-DCIS (%)**	**Total**
Truth			
UDH	13 (81.3)	3 (18.7)	16
LG-DCIS	47 (9.6)	444 (90.4)	491
Total	60	447	507

Accuracy=90.1%, Mahalanobis' distance=2.24234. UDH: Usual ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

	Prediction		
	UDH (%)	**IM-DCIS (%)**	**Total**
Truth			
UDH	14 (87.5)	2 (12.5)	16
IM-DCIS	6 (1.7)	345 (98.3)	351
Total	20	347	367

Accuracy=97.8%, Mahalanobis' distance=4.25055. UDH: Usual ductal hyperplasia, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*

	Prediction		
	UDH (%)	**HG-DCIS (%)**	**Total**
Truth			
UDH	16 (100)	0 (0)	16
HG-DCIS	2 (0.8)	249 (99.2)	251
Total	18	249	267

Accuracy=99.3%, Mahalanobis' distance=5.40486. UDH: Usual ductal hyperplasia, HG-DCIS: High-grade ductal carcinoma *in situ*

	Prediction		
	ADH (%)	**LG-DCIS (%)**	**Total**
Truth			
ADH	109 (77.9)	31 (22.1)	140
LG-DCIS	44 (9.0)	447 (91.0)	491
Total	153	478	631

Accuracy=88.1%, Mahalanobis' distance=2.48288. ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

	Prediction		
	ADH (%)	**IM-DCIS (%)**	**Total**
Truth			
ADH	126 (90.0)	14 (10.0)	140
IM-DCIS	12 (3.4)	339 (96.6)	351
Total	138	353	491

Accuracy=94.7%, Mahalanobis' distance=3.67725. ADH: Atypical ductal hyperplasia, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*

	Prediction		
	ADH (%)	**HG-DCIS (%)**	**Total**
Truth			
ADH	133 (95.00)	7 (5.0)	140
HG-DCIS	6 (2.4)	245 (97.6)	251
Total	139	252	391

Accuracy=96.7%, Mahalanobis' distance=4.32731. ADH: Atypical ductal hyperplasia, HG-DCIS: High-grade ductal carcinoma *in situ*

Contd...

Table 4a: Contd...

	Prediction		
	LG-DCIS (%)	**IM-DCIS (%)**	**Total**
Truth			
LG-DCIS	417 (84.9)	74 (15.1)	491
IM-DCIS	79 (22.5)	272 (77.5)	351
Total	496	346	842

Accuracy=81.8%, Mahalanobis' distance=1.73567. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*

	Prediction		
	LG-DCIS (%)	**HG-DCIS (%)**	**Total**
Truth			
LG-DCIS	485 (98.8)	6 (1.2)	491
HG-DCIS	16 (6.4)	235 (93.6)	251
Total	501	247	742

Accuracy=97.0%, Mahalanobis' distance=4.01048. LG-DCIS: Low-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

	Prediction		
	IM-DCIS (%)	**HG-DCIS (%)**	**Total**
Truth			
IM-DCIS	340 (96.9)	11 (3.1)	351
HG-DCIS	13 (5.2)	238 (94.8)	251
Total	353	249	602

Accuracy=96.0%, Mahalanobis' distance=3.69370. IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

DIN grades, in which the absence of invasion indicates a noncarcinoma lesion, could substantially modify the clinical concept of this disease.

It is customary to perform an initial pathological diagnosis of a mammary lesion via needle biopsy. If the pathological diagnosis is ADH, upstaging to carcinoma occurs in 11–36% of cases based on subsequent resection material.[20-24] Therefore, at minimum an excisional tumor biopsy is advisable following an ADH diagnosis.[20] However, excisional biopsy itself can place considerable stress on the patient. To avoid this, a new pathological diagnostic assessment is under development. Ely *et al*. reported that ≥3 ADH lesions per needle biopsy specimen indicates a higher risk of upstaging, whereas <3 ADH lesions indicates no risk.[24]

On the other hand, the existence of ADH lesions in the surgical margins of resected breast carcinoma material is of clinical concern. Currently, most researchers do not regard the existence of ADH lesions in the surgical margin as a risk factor for recurrence and, therefore, advise against re-operation.[25-28]

As mentioned above, although the pathological diagnosis of IDPLs is not simple, the strict distinction of the lesions in this category is clinically necessary. The features extracted using our algorithm show promising discriminatory power, with accuracies exceeding 80% when using LDA [Table 4b]. In particular, an 88% accuracy rate was achieved for HG-DCIS. This might reflect the fact that pathologists, generally, diagnose HG-DCIS based on nuclear atypia. On the other hand, the accuracy rate for ADH was relatively low except when distinguishing IM-DCIS and HG-DCIS. However, an 88.1% discrimination rate was achieved between ADH and LG-DCIS, which is the most problematic for distinction by visual inspection by pathologists. In the case of SVM analysis, the selection of training data set and prediction data set may be affect the accuracy level. In this analysis, all ROI images were treated as independent data. But ROIs came from the same case may have correlation each other. For checking this, we selected the training data set and prediction data set by case bases, and test ADH and LG-DCIS SVM discrimination for 4 times.

The accuracy results are 71.8% (correct predicted ROIs/prediction data set ROIs = 23/32), 75.6% (68/90), 76.5% (75/98), and 95.9% (163/170), the result cross-validation 88% becomes almost average for case level testing. Invasive breast carcinomas had large intercase and intracase heterogeneity. In the case of intraductal lesions, the nuclei-based heterogeneity was small comparing invasive cases. There was the possibility that results of discrimination keep around 80% accuracy level even using only nuclear features. These data suggests that imaging analysis might provide useful support for this distinction.

Table 4c shows the Mahalanobis distance value of two group centroid values on each group distribution. When this value has the small value, two group data have the close features data values. For example, Mahalanobis distance value between Normal and ADH group (1.98) is smaller than Normal and HG-DCIS (4.37), Normal and ADH nuclei have the near measured feature data values intuitively. For considering the similarity of each group data set, we use the Mahalanobis distance value by magnitude correlation. The Mahalanobis distance between each group centroid showed that the smallest distance value between LG-DCIS and IM-DCIS (1.74) followed by the distance between ADH and normal (1.98). The distance value between ADH and LG-DCIS was larger than these distance, indicating that considerable difference in nucleic features lies between ADH and LG-DCIS. It is difficult to explain what individual difference in nuclear features mean biologically and pathologically. However, it is noteworthy that the large distance lies between ADH and LG-DCIS which are considered to be difficult to differentiate morphologically by pathologist's eyes. This large distance indicates that these two diseases are biologically different and should be differentiated strictly at pathological diagnosis.

Table 4b: Accuracy table

	Normal	UDH (%)	ADH (%)	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)
Normal		97.3	84.8	93.0	97.9	97.6
UDH			87.2	90.1	97.8	99.3
ADH				88.1	94.7	96.7
LG-DCIS					81.8	97.0
IM-DCIS						96.0
HG-DCIS						

UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

Table 4c: Mahalanobis' distance

	Normal	UDH	ADH	LG-DCIS	IM-DCIS	HG-DCIS
Normal		3.88	1.98	3.26	4.13	4.37
UDH			2.25	2.69	4.25	5.40
ADH				2.48	3.68	4.33
LG-DCIS					1.74	4.01
IM-DCIS						3.69
HG-DCIS						

UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

Table 5a: Step-wise linear discriminant analysis: Nuclear size and shape and intranuclear texture features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	147 (86.5)	5 (2.9)	10 (5.9)	8 (4.7)	170
UDH	0 (0)	14 (87.5)	0 (0)	2 (12.5)	16
ADH	16 (11.4)	7 (5.0)	100 (71.4)	17 (12.9)	140
LG-DCIS	18 (3.7)	11 (2.2)	33 (6.7)	429 (87.4)	491
Total	181	37	143	456	817

Total accuracy=84.5%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

Table 5b: Linear kernel SVM discriminant analysis: Nuclear size and shape and intranuclear texture features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	170 (100)	0 (0)	0 (0)	0 (0)	170
UDH	0 (0)	16 (100)	0 (0)	0 (0)	16
ADH	0 (0)	0 (0)	139 (99.3)	1 (0.7)	140
LG-DCIS	0 (0)	0 (0)	1 (0.2)	490 (99.8)	491
Total	170	16	140	491	817

Total accuracy=99.8%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, SVM: Support vector machine

Moreover, the distance value between UDH and normal (3.88) was larger than that of LG-DCIS and normal (3.26), as well as that of ADH and normal. Although UDH is regarded as the disease that has the lowest risk to develop to invasive cancer in IDPLs, it may be possible that UDH has a different character from the other IDPLs from the viewpoint of nucleic features.

The smallest distance value was identified between LG-DCIS and IM-DCIS. When LG-DCIS is diagnosed by pathologists, the major ground of diagnosis is its structural atypia along with the monotonous nuclear shape of individual cells. If the nuclear pleomorphism exists to some extent, but not prominent like HG-DCIS, the lesion is diagnosed as IM-DCIS. Therefore, when the analytic features of over a

Table 5c: Step-wise linear discriminant analysis: Nuclear size and shape features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	100 (58.8)	28 (16.5)	27 (15.9)	15 (8.8)	170
UDH	1 (6.3)	12 (75.0)	2 (12.5)	1 (6.3)	16
ADH	33 (46.2)	17 (12.1)	72 (51.4)	18 (12.9)	140
LG-DCIS	45 (9.2)	88 (17.9)	35 (7.1)	323 (65.8)	491
Total	179	145	136	357	817

Total accuracy=62.1%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

Table 5d: Linear kernel SVM discriminant analysis: Nuclear size and shape features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	118 (69.4)	0 (0)	12 (7.0)	40 (23.6)	170
UDH	0 (0)	9 (56.3)	1 (6.3)	6 (37.4)	16
ADH	23 (16.4)	0 (0)	71 (50.7)	46 (32.9)	140
LG-DCIS	25 (5.1)	0 (0)	10 (2.0)	456 (92.9)	491
Total	166	9	94	548	817

Total accuracy=80.0%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, SVM: Support vector machine

Table 5e: Step-wise linear discriminant analysis: Intranuclear texture features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	147 (86.5)	6 (3.5)	11 (6.5)	6 (3.5)	170
UDH	1 (6.3)	14 (87.5)	1 (6.3)	0 (0)	16
ADH	20 (14.3)	8 (5.7)	90 (64.3)	22 (15.7)	140
LG-DCIS	23 (4.7)	17 (3.5)	37 (7.5)	414 (84.3)	491
Total	191	45	39	442	817

Total accuracy=81.4%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*

Table 5f: Linear kernel SVM discriminant analysis: Intranuclear texture features

	Prediction				
	Normal (%)	UDH (%)	ADH (%)	LG-DCIS (%)	Total
Truth					
Normal	170 (100)	0 (0)	0 (0)	0 (0)	170
UDH	0 (0)	16 (100)	0 (0)	0 (0)	16
ADH	0 (0)	0 (0)	140 (99.3)	0 (0)	140
LG-DCIS	0 (0)	0 (0)	2 (0.4)	489 (99.6)	491
Total	170	16	142	489	817

Total accuracy=99.8%. UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, SVM: Support vector machine

thousand nuclei in each ROI images were averaged, the Mahalanobis distance value between LG-DCIS and IM-DCIS resulted in relatively small. This is supported by the larger distance value between IM-DCIS and HG-DCIS (3.69), in which the nuclear pleomorphism is prominent.

To explore the nuclear characteristics used for pathological diagnosis of intraductal lesions, the 472 imaging analysis features were simply classified as size and shape-related and intranuclear texture-related prior to further analysis. Interestingly, the intranuclear texture was found to be more important than size and shape when discriminating

Table 6a: Step-wise linear discriminant analysis: Nuclear size and shape and intranuclear texture features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	389 (86.1)	84 (17.1)	18 (3.7)	491
IM-DCIS	95 (27.1)	241 (68.7)	15 (4.3)	351
HG-DCIS	14 (5.6)	15 (6.0)	222 (88.4)	251
Total	498	340	255	1093

Total accuracy=78.0%. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

Table 6b: Linear kernel SVM discriminant analysis: Nuclear size and shape and intranuclear texture features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	431 (87.8)	56 (11.4)	4 (0.8)	491
IM-DCIS	98 (27.9)	248 (70.7)	5 (1.4)	351
HG-DCIS	11 (4.4)	5 (2.0)	235 (93.6)	251
Total	540	309	244	1093

Total accuracy=83.6%. LG-DCIS: Low-grade ductal carcinoma *in situ,* IM-DCIS: Intermediate grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*, SVM: Support vector machine

Table 6c: Step-wise linear discriminant analysis: Nuclear size and shape features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	373 (76.0)	95 (18.3)	23 (4.7)	491
IM-DCIS	108 (30.8)	209 (59.5)	34 (9.7)	351
HG-DCIS	25 (10.0)	15 (6.0)	211 (84.1)	251
Total	506	319	268	1093

Total accuracy=72.6%. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

among normal duct, UDH, ADH, and LG-DCIS. On the other hand, nucleus size and shape and intranuclear texture were equally important only when grading DCIS, and therefore, both characteristics are apparently necessary for this process.

In summary, this study has shown that computerized analysis based on a detailed imaging study of nuclei can simulate pathological diagnoses of IDPLs. This study was limited to the analysis of nuclear data and did not consider other tissue data, including size, distribution,

Table 6d: Linear kernel SVM discriminant analysis: Nuclear size and shape features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	428 (87.2)	51 (10.4)	12 (2.4)	491
IM-DCIS	145 (41.3)	248 (50.4)	29 (8.3)	351
HG-DCIS	33 (13.1)	14 (5.6)	204 (81.3)	251
Total	606	313	245	1093

Total accuracy=74.0%. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*, SVM: Support vector machine

Table 6e: Step-wise linear discriminant analysis: Intranuclear texture features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	350 (71.2)	121 (24.6)	20 (4.1)	491
IM-DCIS	97 (27.6)	232 (66.1)	22 (6.3)	351
HG-DCIS	27 (10.8)	20 (8.0)	204 (81.3)	251
Total	474	373	246	1093

Total accuracy=71.9%. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*

Table 6f: Linear kernel SVM discriminant analysis: Intranuclear texture features

	Prediction			
	LG-DCIS (%)	IM-DCIS (%)	HG-DCIS (%)	Total
Truth				
LG-DCIS	428 (87.2)	45 (9.2)	18 (3.6)	491
IM-DCIS	134 (38.2)	198 (56.4)	19 (5.4)	351
HG-DCIS	21 (8.4)	16 (6.4)	214 (85.2)	251
Total	583	259	251	1093

Total accuracy=76.9%. LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, HG-DCIS: High-grade ductal carcinoma *in situ*, SVM: Support vector machine

and presence of necrosis. Although this method is not a substitute for visual pathological analysis, the high accuracy rates of the reported methods suggest that it provides practical and useful support for pathological diagnosis, and, therefore, further studies are expected.

Financial Support and Sponsorship

Nil.

Conflicts of Interests

The authors declare that they have no competing interests.

Table 7: Contribution level of each feature via SVM analysis

	Contribution (%)	Accumulated (%)
GLCM contrast average	18.6	18.6
GLCM contrast median	12.7	31.3
Contour line complexity variance	8.5	39.8
IDG7 average	8.3	48.1
Nuclear texture complexity average	5.9	54.0
GLCM homogeneity variance	5.3	59.3
GLCM homogeneity average	5.2	64.5
GLCM angular 2nd median	4.8	69.3
GLCM entropy variance	3.8	73.1
Long axis median	3.5	76.6
IDG7 variance	3.4	80.0
Long axis variance	3.2	83.2
IDG2 variance	2.8	86.0
Nuclear texture complexity variance (80% based)	2.5	88.5
IDG4 median	2.2	90.7
Short axis variance	2.2	92.9
IDG4 variance	2.1	95.0
Contour line complexity median	1.9	96.9
Roundness variance	1.6	98.5
IDG2 variance (80% based)	1.1	99.6
GLCM angular 2nd variance	0.3	99.9
GLCM entropy median	0.1	100.0

GLCM: Gray level co-occurrence matrix, IDG: Integrated diffusion gradient, SVM: Support vector machine

REFERENCES

1. Wells WA, Carney PA, Eliassen MS, Tosteson AN, Greenberg ER. Statewide study of diagnostic agreement in breast pathology. J Natl Cancer Inst 1998;90:142-5.
2. Palli D, Galli M, Bianchi S, Bussolati G, Di Palma S, Eusebi V, *et al.* Reproducibility of histological diagnosis of breast lesions: Results of a panel in Italy. Eur J Cancer 1996;32A: 603-7.
3. Jain RK, Mehta R, Dimitrov R, Larsson LG, Musto PM, Hodges KB, *et al.* Atypical ductal hyperplasia: Interobserver and intraobserver variability. Mod Pathol 2011;24:917-23.
4. Tavassoli FA. Ductal intraepithelial neoplasia of the breast. Virchows Arch 2001;438:221-7.
5. Elston CW, Sloane JP, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, *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;36:1769-72.
6. Page DL, Vander Zwaag R, Rogers LW, Williams LT, Walker WE, Hartmann WH. Relation between component parts of fibrocystic disease complex and breast cancer. J Natl Cancer Inst 1978;61:1055-63.
7. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. Cancer 1985;55:2698-708.
8. 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;65:518-29.
9. Page DL, Rogers LW. Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia. Hum Pathol 1992;23:1095-7.
10. Tavassoli FA. Mammary intraepithelial neoplasia: A translational classification system for the intraductal epithelial proliferations. Breast J 1997;3:48-58.
11. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. N Engl J Med 1985;312:146-51.
12. Saito A, Cosatto E, Kiyuna T, Sakamoto M. Dawn of the Digital Diagnosis Assisting System, Can it Open a New Age for Pathology? Proc SPIE 2013; 8676:867602.
13. Cosatto E, Miller M, Graf HP, Meyer JS. Grading Nuclear Pleomorphism on Histological Micrographs. Pattern Recognition, 2008. ICPR 2008. 19th International Conference on; 2008. p. 1-4.
14. Cosatto E, Laquerre PF, Malon C, Graf HP, Saito A, Kiyuna T, *et al.* Automated Gastric Cancer Diagnosis on H and E-stained Sections; training a classifier on large scale with multiple instance machine learning. Proc SPIE 2013;8676 :867605.
15. Lakhani SR, Eiilis IO, Schnitt SJ. WHO Classification of Tumours of the Breast. 4th ed. International Agency for Research on Cancer; 2012. p. 81-94.
16. Nguyen CV, Albarracin CT, Whitman GJ, Lopez A, Sneige N. Atypical ductal hyperplasia in directional vacuum-assisted biopsy of breast microcalcifications: Considerations for surgical excision. Ann Surg Oncol 2011;18:752-61.
17. Schnitt SJ, Connolly JL, Tavassoli FA, Fechner RE, Kempson RL, Gelman R, *et al.* Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am J Surg Pathol 1992;16:1133-43.
18. 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;4:230-7.
19. Rosai J. Borderline epithelial lesions of the breast. Am J Surg Pathol 1991;15:209-21.
20. McGhan LJ, Pockaj BA, Wasif N, Giurescu ME, McCullough AE, Gray RJ. Atypical ductal hyperplasia on core biopsy: An automatic trigger for excisional biopsy? Ann Surg Oncol 2012;19:3264-9.
21. Bendifallah S, Defert S, Chabbert-Buffet N, Maurin N, Chopier J, Antoine M, *et al.* Scoring to predict the possibility of upgrades to malignancy in atypical ductal hyperplasia diagnosed by an 11-gauge vacuum-assisted biopsy device: An external validation study. Eur J Cancer 2012;48:30-6.
22. Polat AK, Kanbour-Shakir A, Andacoglu O, Polat AV, Johnson R, Bonaventura M, *et al.* Atypical hyperplasia on core biopsy: Is further surgery needed? Am J Med Sci 2012;344:28-31.
23. Winchester DJ, Bernstein JR, Jeske JM, Nicholson MH, Hahn EA, Goldschmidt RA, *et al.* Upstaging of atypical ductal hyperplasia after vacuum-assisted 11-gauge stereotactic core needle biopsy. Arch Surg 2003;138:619-22.
24. Ely KA, Carter BA, Jensen RA, Simpson JF, Page DL. Core biopsy of the breast with atypical ductal hyperplasia: A probabilistic approach to reporting. Am J Surg Pathol 2001;25:1017-21.
25. Greene T, Tartter PI, Smith SR, Estabrook A. The significance of surgical margins for patients with atypical ductal hyperplasia. Am J Surg 2006;192:499-501.
26. Arora S, Menes TS, Moung C, Nagi C, Bleiweiss I, Jaffer S. Atypical ductal hyperplasia at margin of breast biopsy – is re-excision indicated? Ann Surg Oncol 2008;15:843-7.
27. Baker JL, Hasteh F, Blair SL. Atypical Ductal hyperplasia at the margin of lumpectomy performed for early stage breast cancer: Is there enough evidence to formulate guidelines? Int J Surg Oncol 2012;2012:297832.
28. Li S, Liu J, Yang Y, Zeng Y, Deng H, Jia H, *et al.* Impact of atypical hyperplasia at margins of breast-conserving surgery on the recurrence of breast cancer. J Cancer Res Clin Oncol 2014;140:599-605.

Supplementary Table 1: Step-wise discriminant function (6 group-based) coefficient values

	0 versus 1	0 versus 2	0 versus 3	0 versus 4	0 versus 5	1 versus 2	1 versus 3	1 versus 4	1 versus 5	2 versus 3	2 versus 4	2 versus 5	3 versus 4	3 versus 5	4 versus 5
	4.13728	**2.95486**	**3.52221**	**3.88048**	**4.76612**	**4.1185**	**3.94527**	**4.26791**	**5.71601**	**3.33386**	**3.54664**	**4.91772**	**1.95562**	**4.32925**	**4.08723**
Contour line length variance	−2.45595	3.90460	17.93029	34.47757	70.59923	6.36060	20.38624	36.93352	73.05518	14.02564	30.57292	66.69458	16.54728	52.66900	36.12166
Long axis variance	24.43368	14.17900	8.15686	4.31150	−12.02098	−10.25452	−16.27681	−20.12218	−36.45465	−6.02229	−9.86766	−26.20013	−3.84537	−20.17800	−16.33247
IDG26 variance	−2.99068	−0.33181	−12.42668	−5.00465	−41.57119	2.65887	−9.43600	−2.01397	−38.58051	−12.09487	−4.67284	−41.23938	7.42203	−29.14500	−36.56654
Area variance (80% based)	−10.06658	−9.17380	−5.80230	−6.96151	−8.92086	0.89281	4.26428	3.10507	1.14572	3.37148	2.21226	0.25291	−1.15921	−3.11860	−1.95935
IDG20 SD	−0.35924	1.13260	0.67838	0.31308	−0.10298	1.49186	1.03761	0.67232	0.25626	−0.45424	−0.81954	−1.23560	−0.36530	−0.78136	−0.41606
Short axis median	−0.81979	−0.83140	0.32665	3.30409	3.83763	−0.01161	1.14644	4.12388	4.65742	1.15805	4.13549	4.66902	2.97744	3.51100	0.53354
IDG17 average	−1.84136	−2.27430	−1.44537	1.14898	2.84772	−0.43296	0.39599	2.99034	4.68908	0.82895	3.42329	5.12204	2.59434	4.29310	1.69875
GLCM contrast SD	10.97419	0.35968	−6.74873	−7.75936	−6.77617	−10.61451	−17.72293	−18.73355	−17.75036	−7.10842	−8.11904	−7.13585	−1.01063	−0.02744	0.98319
Density SD	−4.62533	0.03327	−0.49123	−0.32324	1.40839	4.65860	4.13410	4.30209	6.03372	−0.52449	−0.35651	1.37512	0.16798	1.89960	1.73163
GLCM homogeneity SD (80% based)	−1.25578	−1.66800	−1.25044	−2.97417	−2.52123	−0.41226	0.00534	−1.71840	−1.26546	0.41759	−1.30614	−0.85320	−1.72373	−1.27080	0.45294
Contour line complexity mode	−0.43924	0.00733	0.23295	−0.38108	0.15111	0.44657	0.67220	0.05816	0.59035	0.22562	−0.38841	0.14378	−0.61403	−0.08184	0.53219
Density 70% tile	8.44891	0.83034	−0.27774	−0.00884	−1.72679	−7.61857	−8.72665	−8.45775	−10.17570	−1.10808	−0.83918	−2.55713	0.26890	−1.44910	−1.71795
GLCM contrast 90% tile	2.51385	3.37090	4.98107	2.82410	1.27680	0.85709	2.46722	0.31025	−1.23705	1.61013	−0.54684	−2.09414	−2.15697	−3.70430	−1.54731
IDG22 average (80% based)	−3.26739	20.43300	19.59399	16.96300	13.23329	23.70061	22.86138	20.23039	16.50069	−0.83923	−3.47022	−7.19992	−2.63100	−6.36070	−3.72970
Ellipsoidal ratio SD (80% based)	−0.61951	−1.24260	−0.47980	−1.14254	0.19152	−0.62311	0.13971	−0.52303	0.81102	0.76282	0.10008	1.43413	−0.66274	0.67131	1.33406
IDG21 SD	−3.96381	−0.57920	−1.89173	−1.92866	−2.85580	3.38461	2.07209	2.03516	1.10801	−1.31252	−1.34945	−2.27660	−0.03693	−0.96408	−0.92715
Grad SD	9.08664	3.61500	3.26894	2.56590	8.84527	−5.47167	−5.81770	−6.52074	−0.24137	−0.34603	−1.04907	5.23030	−0.70304	5.57630	6.27937
Short axis SD	−0.63408	0.59537	1.59699	0.03651	4.10539	1.22945	2.23107	0.67060	4.73947	1.00162	−0.55885	3.51002	−1.56047	2.50840	4.06887
GLCM homogeneity 10% tile	0.67110	−1.35250	2.66831	0.23901	0.51067	−2.02357	1.99721	−0.43208	−0.16042	4.02078	1.59148	1.86315	−2.42930	−2.15760	0.27166
GLCM entropy 10% tile	−5.11912	−3.29270	1.38613	−1.60646	−1.49876	1.82644	6.50525	3.51266	3.62036	4.67881	1.68622	1.79392	−2.99258	−2.88490	0.10770
IDG6 30% tile	2.88927	0.42665	0.61095	−2.80860	−7.47509	−2.46262	−2.27832	−5.69787	−10.36436	0.18430	−3.23525	−7.90174	−3.41955	−8.08600	−4.66649
IDG2 median	−6.50187	−3.06410	−2.82203	−2.90891	−0.40886	3.43772	3.67984	3.59296	6.09301	0.24212	0.15524	2.65529	−0.08688	2.41320	2.50005
Nucleus texture complexity SD (80% based)	3.67331	2.30680	1.36598	1.92199	−0.55359	−1.36652	−2.30733	−1.75133	−4.22691	−0.94082	−0.38481	−2.86039	0.55601	−1.91960	−2.47558
IDG2 90% tile	1.90388	−0.96480	1.24658	1.88233	2.58192	−2.86868	−0.65730	−0.02155	0.67805	2.21137	2.84713	3.54672	0.63575	1.33530	0.69959
IDG22 variance (80% based)	10.29505	9.53660	6.62849	6.80900	11.10975	−0.75848	−3.66655	−3.48604	0.81470	−2.90807	−2.72756	1.57319	0.18051	4.48130	4.30075
ISG14 variance (80% based)	0.15548	−1.74040	−0.70318	0.77219	−0.74140	−1.89584	−0.85866	0.61671	−0.89688	1.03719	2.51255	0.99896	1.47537	−0.03823	−1.51360
IDG12 10% tile	−0.52918	−1.99490	−2.42524	−1.57572	−0.48215	−1.46570	−1.89605	−1.04654	0.04704	−0.43035	0.41916	1.51274	0.84951	1.94310	1.09357
IDG22 SD (80% based)	−13.85679	−9.92630	−8.37700	−7.70852	−11.36982	3.93049	5.47978	6.14827	2.48697	1.54929	2.21778	−1.44352	0.66849	−2.99280	−3.66130
IDG12 30% tile	−0.41849	2.21210	−0.13480	1.58529	−0.90775	2.63064	0.28370	2.00378	−0.48925	−2.34694	−0.62686	−3.11989	1.72008	−0.77295	−2.49303
IDG25 SD	−3.87064	−6.01200	−3.39737	−0.80152	−1.98267	−2.14131	0.47327	3.06912	1.88797	2.61458	5.21043	4.02928	2.59585	1.41470	−1.18115
GLCM angular 2nd 90% tile	−11.16926	−4.80560	−4.62633	−5.83675	−6.42173	6.36363	6.54293	5.33251	4.74754	0.17931	−1.03112	−1.61609	−1.21042	−1.79540	−0.58498
IDG21 variance	3.70138	0.00009	2.17455	2.51966	3.02900	−3.70128	−1.52683	−1.18172	−0.67238	2.17445	2.51956	3.02891	0.34511	0.85445	0.50934
Long axis variance (80% based)	−0.44984	0.38347	0.01026	5.24942	7.34405	0.83330	0.46009	5.69926	7.79388	−0.37321	4.86596	6.96058	5.23917	7.33380	2.09462
Contour line length variance (80% based)	−0.22958	−4.23080	−1.18044	−13.23308	−3.63767	−4.00126	−0.95086	−13.00350	−3.40809	3.05040	−9.00224	0.59317	−12.05264	−2.45720	9.59541

Contd...

Supplementary Table 1: Contd...

	0 versus 1	0 versus 2	0 versus 3	0 versus 4	0 versus 5	1 versus 2	1 versus 3	1 versus 4	1 versus 5	2 versus 3	2 versus 4	2 versus 5	3 versus 4	3 versus 5	4 versus 5
	4.13728	**2.95486**	**3.52221**	**3.88048**	**4.76612**	**4.1185**	**3.94527**	**4.26791**	**5.71601**	**3.33386**	**3.54664**	**4.91772**	**1.95562**	**4.32925**	**4.08723**
Contour line complexity 30% tile	0.43644	0.35515	−0.53079	−0.91951	−0.16456	−0.08129	−0.96723	−1.35595	−0.60100	−0.88594	−1.27466	−0.51972	−0.38871	0.36623	0.75494
Contour line complexity SD	3.48830	−1.68990	−1.83937	−1.79029	−9.90466	−5.17824	−5.32766	−5.27858	−13.39296	−0.14943	−0.10034	−8.21472	0.04908	−8.06530	−8.11438
Contour line complexity SD (80% based)	−0.14065	1.47270	1.24789	1.29205	10.39673	1.61340	1.38854	1.43271	10.53738	−0.22485	−0.18069	8.92399	0.04417	9.14880	9.10467
Density 10% tile	−0.96943	2.35320	2.62938	4.26228	4.57133	3.32259	3.59881	5.23171	5.54076	0.27621	1.90912	2.21817	1.63290	1.94200	0.30905
Density 30% tile	−3.64530	−3.04870	−4.35690	−6.45799	−4.37961	0.59665	−0.71160	−2.81269	−0.73432	−1.30825	−3.40934	−1.33096	−2.10109	−0.02271	2.07838
GLCM contrast mod	−1.39641	−1.16910	−0.75226	−0.40498	−0.94783	0.22735	0.64416	0.99144	0.44858	0.41680	0.76408	0.22123	0.34728	−0.19558	−0.54286
IDG1 mod (80% based)	−0.31355	−0.74023	−0.37708	−0.17387	0.41487	−0.42668	−0.06353	0.13969	0.72843	0.36315	0.56636	1.15511	0.20322	0.79196	0.58874
GLCM entropy variance	−12.92171	−3.81570	−0.32064	−0.81774	−7.40643	9.10603	12.60107	12.10397	5.51528	3.49504	2.99794	−3.59075	−0.49710	−7.08580	−6.58869
GLCM angular 2nd variance (80% based)	1.16789	−0.75830	0.09215	−2.04556	−0.72528	−1.92619	−1.07574	−3.21345	−1.89317	0.85045	−1.28726	0.03303	−2.13771	−0.81743	1.32029
Long axis SD	−15.00882	−11.42200	−9.14040	−9.36411	−18.49248	3.58670	5.86842	5.64471	−3.48366	2.28172	2.05802	−7.07035	−0.22371	−9.35210	−9.12837
IDG26 SD	5.00569	5.21000	12.26505	3.62544	26.15301	0.20430	7.25936	−1.38025	21.14733	7.05506	−1.58454	20.94303	−8.63961	13.88800	22.52757
IDG1 SD (80% based)	12.13854	7.70830	−0.27953	2.52418	−12.27263	−4.43026	−12.41807	−9.61436	−24.41117	−7.98781	−5.18410	−19.98091	2.80371	−11.99300	−14.79681
Contour line length 70% tile	−7.27578	−6.10100	−7.69497	−5.77138	−7.48826	1.17477	−0.41919	1.50441	−0.21248	−1.59396	0.32964	−1.38724	1.92360	0.20672	−1.71688
Roundness median	−0.32965	0.10934	−2.41092	−0.01602	−0.98204	0.43899	−2.08127	0.31363	−0.65239	−2.52027	−0.12536	−1.09139	2.39491	1.42890	−0.96602
Ellipsoidal ratio 30% tile	−1.36816	−3.71760	−0.19303	−1.40973	1.10127	−2.34946	1.17513	−0.04157	2.46943	3.52460	2.30790	4.81889	−1.21670	1.29430	2.51100
Long axis average (80% based)	−7.26713	0.49381	3.28564	4.99634	18.81654	7.76094	10.55277	12.26347	26.08367	2.79183	4.50253	18.32273	1.71070	15.53100	13.82020
IDG22 variance	2.77874	2.74860	2.95507	1.48810	−2.02308	−0.03018	0.17634	−1.29064	−4.80181	0.20651	−1.26046	−4.77163	−1.46697	−4.97810	−3.51118
Nucleus arrayment level 50% tile	−0.53787	−0.25212	−0.08486	−0.50036	0.67438	0.28575	0.45301	0.03751	1.21225	0.16726	−0.24824	0.92650	−0.41550	0.75924	1.17474
Nucleus arrayment level 70% tile	0.56425	0.32856	0.13287	0.85142	−0.94208	−0.23569	−0.43138	0.28717	−1.50633	−0.19568	0.52286	−1.27063	0.71854	−1.07500	−1.79349
IDG11 mod (80% based)	0.87931	−0.08627	0.33103	0.32860	−0.54182	−0.96559	−0.54828	−0.55071	−1.42113	0.41731	0.41488	−0.45554	−0.00243	−0.87285	−0.87042
Nucleus arrayment level SD	−0.00599	−0.47191	0.21993	−0.09366	1.55954	−0.46593	0.22591	−0.08767	1.56552	0.69184	0.37825	2.03145	−0.31359	1.33960	1.65320
IDG15 SD	−0.56547	0.04998	−1.00986	−0.74221	−0.72104	0.61545	−0.44439	−0.17675	−0.15557	−1.05984	−0.79220	−0.77102	0.26765	0.28882	0.02118
GLCM angular 2nd variance	6.08093	3.24780	0.42180	3.78794	4.31861	−2.83313	−5.65913	−2.29299	−1.76232	−2.82600	0.54013	1.07081	3.36614	3.89680	0.53068
GLSM angular 2nd SD	−1.60532	−2.05200	−0.41323	−2.78629	−2.97491	−0.44668	1.19209	−1.18098	−1.36960	1.63877	−0.73430	−0.92292	−2.37306	−2.56170	−0.18862
Roundness average (80% based)	1.32855	4.42800	3.44396	2.55956	3.93802	3.09948	2.11542	1.23102	2.60948	−0.98406	−1.86846	−0.49000	−0.88440	0.49406	1.37846
GLCM contrast variance	−10.42569	−2.22840	5.38329	6.86778	7.35482	8.19725	15.80899	17.29347	17.78051	7.61174	9.09622	9.58327	1.48448	1.97150	0.48704
GLCM contrast variance (80% based)	2.48528	0.97243	−3.89470	−4.85080	−2.74092	−1.51285	−6.37998	−7.33607	−5.22619	−4.86713	−5.82322	−3.71334	−0.95610	1.15380	2.10988
GLCM contrast SD (80% based)	−1.76952	0.10490	3.66781	5.88013	3.18267	1.87442	5.43733	7.64965	4.95219	3.56291	5.77523	3.07777	2.21233	−0.48514	−2.69747
IDG13 SD (80% based)	1.01523	0.12061	0.96890	−0.08321	0.50770	−0.89462	−0.04633	−1.09843	−0.50753	0.84829	−0.20382	0.38709	−1.05211	−0.46120	0.59091
IDG17 average (80% based)	2.33720	0.11703	2.44958	0.65271	−2.59262	−2.22016	0.11239	−1.68449	−4.92981	2.33255	0.53567	−2.70965	−1.79687	−5.04220	−3.24532
IDG23 variance	−0.38115	−0.59816	−0.84980	−1.12236	2.76502	−0.21701	−0.46864	−0.74121	3.14617	−0.25164	−0.52420	3.36318	−0.27256	3.61480	3.88738
Area size average	−22.75324	−5.41310	−28.47442	−61.35417	−70.21588	17.34014	−5.72119	−38.60093	−47.46264	−23.06133	−55.94107	−64.80279	−32.87974	−41.74100	−8.86171
IDG24 SD (80% based)	−0.58727	0.22878	−0.31580	−1.07611	−2.91232	0.81605	0.27147	−0.48884	−2.32506	−0.54458	−1.30489	−3.14111	−0.76031	−2.59650	−1.83622

Contd...

Supplementary Table 1: Contd...

	0 versus 1	0 versus 2	0 versus 3	0 versus 4	0 versus 5	1 versus 2	1 versus 3	1 versus 4	1 versus 5	2 versus 3	2 versus 4	2 versus 5	3 versus 4	3 versus 5	4 versus 5
	4.13728	**2.95486**	**3.52221**	**3.88048**	**4.76612**	**4.1185**	**3.94527**	**4.26791**	**5.71601**	**3.33386**	**3.54664**	**4.91772**	**1.95562**	**4.32925**	**4.08723**
ISG1 average (80% based)	29.10597	10.73500	25.88919	51.54611	51.46470	−18.37071	−3.21677	22.44014	22.35873	15.15393	40.81084	40.72943	25.65691	25.57600	−0.08141
IDG5 10% tile	1.16214	1.20220	1.97845	2.75703	2.66372	0.04006	0.81630	1.59489	1.50158	0.77625	1.55483	1.46152	0.77859	0.68528	−0.09331
ISG22 average	10.02033	−19.15800	−17.05032	−11.99863	−11.94108	−29.17826	−27.07065	−22.01896	−21.96142	2.10761	7.15930	7.21684	5.05169	5.10920	0.05755
IDG2 30% tile	0.19910	2.17080	−0.39275	−0.35352	−0.49631	1.97171	−0.59184	−0.55262	−0.69541	−2.56355	−2.52433	−2.66712	0.03923	−0.10357	−0.14280
ISG4 90% tile	−0.77418	1.27180	−0.17212	0.50446	1.49731	2.04594	0.60206	1.27864	2.27148	−1.44388	−0.76729	0.22555	0.67658	1.66940	0.99284
IDG6 average	−2.87781	−4.76460	−1.22855	−5.44260	−4.77888	−1.88675	1.64926	−2.56479	−1.90107	3.53601	−0.67804	−0.01432	−4.21405	−3.55030	0.66372
GLCM angular 2nd SD (80% based)	7.14720	3.53830	2.50893	5.58431	5.07808	−3.60888	−4.63826	−1.56289	−2.06912	−1.02939	2.04599	1.53976	3.07537	2.56920	−0.50622
ISG12 90% tile	1.10518	0.52241	0.52764	0.66042	1.46261	−0.58277	−0.57754	−0.44475	0.35743	0.00523	0.13802	0.94020	0.13279	0.93497	0.80219
ISG25 SD (80% based)	−0.22294	1.52700	1.22201	−1.55035	1.96462	1.74993	1.44495	−1.32741	2.18757	−0.30498	−3.07734	0.43764	−2.77236	0.74262	3.51498
Nuclear texture complexity median	−1.92996	−0.17814	−0.16591	−1.36916	−0.16574	1.75182	1.76405	0.56080	1.76422	0.01223	−1.19102	0.01240	−1.20325	0.00017	1.20342
IDG25 mod	−0.17905	0.09568	0.61251	0.48644	−0.04049	0.27473	0.79157	0.66549	0.13856	0.51683	0.39076	−0.13617	−0.12607	−0.65300	−0.52693

IDG: Integrated diffusion gradient, IDG1: Ratio of nucleus area size and rectangle box area (long × short axis), IDG2: Ratio of nucleus 3D volume to cuboid volume, IDG3-8: Total volume over 6 threshold intensity level, IDG9-14: Increased volume over each threshold intensity level, IDG15-20: Counts for each threshold intensity level cluster, IDG21-26: Image fractal dimensions for each threshold intensity level, GLCM: Gray level co-occurrence matrix, SD: Standard deviation

Supplementary Table 2: Step-wise discriminant function (2 group-based) coefficient values

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	**ADH**	**LG-DCIS**	**IM-DCIS**	**LG-DCIS**	**ADH**	**LG-DCIS**	**IM-DCIS**	**LG-DCIS**	**LG-DCIS**	**IM-DCIS**	**LG-DCIS**	**IM-DCIS**	**LG-DCIS**	**LG-DCIS**
Area size variance											−13.27223	−18.29346			−7.90478
Area size SD			57.83725		−16.19010										
Area size mod											−1.04658				
Area size 10% tile					2.11399										
Area size 90% tile														−4.09205	
Area size variance (80% based)					−15.71332									−8.24592	
Area size SD (80% based)									−22.22490						
Contour line length variance			19.28492	21.75449	46.74395						23.27931	62.76269		41.19523	32.11084
Contour line length SD									43.91830						
Contour line length mod					0.82866									0.89140	
Contour line length 30% tile													2.96827		
Contour line length variance (80% based)													−3.63387		
Roundness variance													−1.18852		
Roundness SD												2.09947			
Roundness median										−1.69265					
Roundness mod				0.52502											
Roundness 10% tile				2.64550											
Roundness 30% tile						1.48289									
Roundness 70% tile	3.74410									−2.23177					

Contd...

Supplementary Table 2: Contd...

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	ADH	LG-DCIS	IM-DCIS	LG-DCIS	ADH	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS
Roundness average (80% based)		2.46756													
Roundness variance (80% based)															−5.61577
Roundness SD (80% based)													1.74487		6.46773
Roundness mod (80% based)														0.75071	
Long axis variance												−19.39143		−10.37206	
Long axis SD					−14.11932				−18.65840					−8.83023	−15.61892
Long axis 90% tile			−3.69966								−4.71198				
Long axis average (80% based)	6.88350													19.66377	
Long axis variance (80% tile)													5.24019		
Long axis mod (80% based)											−0.72179				
Short axis SD															5.46576
Short axis median					4.06523							3.18513	8.10471		
Short axis mod											−2.31099			0.76368	
Short axis 90% tile				−6.23716				−3.94080			−3.85260				
Short axis variance (80% based)					10.23340										
Short axis SD (80% based)					−13.01599										
Ellipsoidal ratio variance					−20.29103										
Ellipsoidal ratio SD					21.00307										
Ellipsoidal ratio median														4.77563	
Ellipsoidal ratio mod	−1.97460														
Ellipsoidal ratio 10%				−1.53372									−0.95176		
Ellipsoidal ratio 70%										2.47981					
Ellipsoidal ratio 90%		−1.16236													
Ellipsoidal ratio SD (80% based)													−1.46922		
Contour line complexity average							−2.07810							−5.75163	
Contour line complexity variance	1.70170				−2.40185										
Contour line complexity SD						−1.87906		−8.40450							
Contour line complexity mod													−0.41150		
Contour line complexity 10% tile		−0.57854							−3.32440		−1.48855				−1.34079
Contour line complexity 70% tile														6.52459	4.18454
Contour line complexity variance (80% based)					4.01273										
Contour line complexity mod (80% based)	1.22120			1.20630									−0.54698		
GLCM angular 2nd variance												1.78125			
GLCM angular 2nd SD								−4.79630							
GLCM angular 2nd mod									−1.12980			−1.17481			
GLCM angular 2nd 10% tile											2.85451				
GLCM angular 2nd 30% tile								1.79580						−0.81772	

Contd...

Supplementary Table 2: Contd...

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	ADH	LG-DCIS	IM-DCIS	LG-DCIS	ADH	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS
GLCM angular 2nd 70% tile											−4.26201				
GLCM angular 2nd SD (80% based)														3.09492	
GLCM contrast variance	−13.24720							14.07790	12.22010						0.89419
GLCM contrast SD	15.95360			1.82566				−16.70180	−14.02120	0.47060					
GLCM contrast median					−11.37614		1.25481	5.00270							−2.03881
GLCM contrast mod					−5.79427										
GLCM contrast 10% tile						1.21010									
GLCM contrast 30% tile														−0.75160	
GLCM contrast 70% tile													1.02559		
GLCM contrast 90% tile					5.73167								−0.84563		
GLCM contrast average (80% based)					10.27625										
GLCM contrast variance (80% based)			−3.76062		−2.85106									2.53430	
GLCM contrast SD (80% based)		0.96029	6.29637											−2.36939	
GLCM contrast mod (80% based)					3.72980										
GLCM homogeneity variance											9.60843			1.66867	
GLCM homogeneity SD							−2.59484			−1.80275	−9.00192				1.61079
GLCM homogeneity median			−2.24356												
GLCM homogeneity 10% tile			4.30371												
GLCM homogeneity variance (80% based)														−3.52999	
GLCM homogeneity SD (80% based)							−3.00048					−1.94652			
GLCM entropy variance								4.36910							
GLCM entropy SD							4.41767			0.98473					
GLCM entropy 10% tile	−2.85230														
GLCM entropy 70% tile			1.85016					4.52600							
GLCM entropy variance (80% based)															−1.21131
GLCM entropy mod (80% based)									1.41900						
IDG1 average														−16.70545	
IDG1 SD			−69.05414	−12.27927				10.98320							
IDG1 median													−4.95015		
IDG1 mod															0.66844
IDG1 10% tile											3.55622				
IDG1 90% tile		−3.33622													
IDG1 SD (80% based)												−19.10310			−9.41216
IDG1 mod (80% based)											1.04149			1.21974	

Contd...

Supplementary Table 2: Contd...

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	ADH	LG-DCIS	IM-DCIS	LG-DCIS	ADH	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS
IDG2 variance				−3.37270						1.93112					
IDG2 SD		−1.45998										2.03789			
IDG2 median	−5.15220								6.95500			3.19972			2.13072
IDG2 mod											0.89989	1.44415		1.04044	
IDG2 10% tile															1.26953
IDG2 30% tile				4.92603								−2.79643			
IDG2 70% tile			−3.46459	−10.98754						−2.13270				6.86090	
IDG2 variance (80% based)				6.56487											
ISG2 SD (80% based)											1.84459				
IDG2 mod (80% based)						1.79717				−0.67010	−1.52524				
IDG4 variance					2.63881										
IDG4 mod					0.61219										
IDG4 10% tile										1.64473		1.81726			
IDG4 average (80% tile)			3.12109								−2.60067				
IDG4 variance (80% based)													−0.51994		
IDG4 mod (80% based)			−0.62008								1.11057				
IDG5 mod															−0.76294
IDG5 70% tile	2.48040														
IDG5 variance (80% based)		1.00729						−4.79370							7.94161
ISG5 SD (80% based)															−7.43929
IDG6 10% tile										1.43176					
IDG7 median							4.69347								
IDG7 30% tile							7.34502								
IDG7 variance (80% based)			−1.87259												
IDG11 median					−2.12799		−4.08882								
IDG11 mod															−0.79547
IDG11 10% tile						−0.93649									
IDG11 70% tile														−2.58924	
IDG11 mod (80% based)									−1.42280	0.54727	1.20923			−0.77511	
IDG12 median							5.59187					−2.33739			
IDG12 10% tile					−2.94088				−3.26240						
IDG12 average (80% based)							−12.47567								
IDG12 variance (80% based)										2.28270					
IDG12 SD (80% based)			3.91714									3.71125	−1.22339		
IDG13 variance					−2.77191										
IDG13 mod												−0.75206		−0.55613	
IDG13 10% tile	−1.23840														
IDG13 30% tile													0.55399		
IDG13 90% tile									2.12380	0.74019		1.80810	−0.77501		1.85487
IDG13 average (80% based)											−2.99129				
IDG13 variance (80% based)												−2.66794			

Contd...

Supplementary Table 2: Contd...

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	ADH	LG-DCIS	IM-DCIS	LG-DCIS	ADH	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS
IDG13 SD (80% based)		−0.54989													
IDG14 30% tile								−6.48350							
IDG14 90% tile				2.48490											
IDG14 variance (80% based)													1.65449		
IDG14 mod (80% based)				1.03274											
IDG15 SD										−0.83170					
IDG15 SD (80% based)							−0.73199					3.30676			
IDG16 median							−0.75284	−1.13520							
IDG16 70% tile		−0.70002	−0.82886				−0.85725								
IDG16 average (80% based)															1.15853
IDG16 SD (80% based)									4.39810						
IDG17 average											3.57911				
IDG17 SD								2.35160		1.42195					
IDG17 90% tile									−2.31810				0.36460		
IDG17 average (80% based)				1.50247											−2.45877
ISG18 SD				−1.27941											
IDG18 SD (80% based)					−2.04079										
IDG19 SD			1.10032				2.22843								
IDG19 90% tile													−0.34184		0.60636
IDG20 SD													−0.38597	−1.27317	
IDG21 variance	6.10210										1.34703				
IDG21 SD	−7.90870			−1.17533	−1.14739										
IDG21 70% tile				0.63598											
IDG21 90% tile													0.42218		
IDG21 average (80% based)			−0.61393											1.83399	
IDG22 variance															−2.10230
IDG22 10% tile												2.06123			
IDG22 30% tile				4.63543											
IDG22 70% tile														−3.83645	
IDG22 variance (80% based)				2.22221											
IDG23 variance			−1.85223	−2.76485						−0.77926					
IDG23 SD											−2.34302				4.07129
IDG23 mod						1.23875					−0.64969				
IDG23 variance (80% based)													−0.95305		
IDG24 variance															−2.81879
IDG24 mod													−0.37591		
IDG24 10% tile			−1.95083	−1.10981				−4.94040							
IDG24 90% tile		−2.07756						−7.96880							
IDG24 mod (80% based)															0.62539
IDG25 variance			−12.00935												

Contd...

Supplementary Table 2: Contd...

	Normal					UDH				ADH			LG-DCIS		IM-DCIS
	UDH	ADH	LG-DCIS	IM-DCIS	LG-DCIS	ADH	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	IM-DCIS	LG-DCIS	LG-DCIS
IDG25 SD			9.09701												
IDG25 variance (80% based)															3.76646
IDG26 variance												−8.94080		−12.96831	−12.28467
IDG26 median	−6.90840														
IDG26 mod				1.27243											
IDG26 30% tile													−4.23981	−6.35785	
IDG26 70% tile				−3.39182											
IDG26 90% tile		5.95983											−2.20614		
IDG26 variance (80% based)					−4.00060				14.93260						
IDG26 SD (80% based)									−19.01620						
IDG26 mod (80% based)								1.47740							
Nuclear texture complexity variance									15.83050			8.18379		4.20196	−1.79967
Nuclear texture complexity SD		1.44804							−23.01030			−12.01935		−6.62293	
Nuclear texture complexity median														−1.07102	
Nuclear texture complexity variance (80% based)			−4.05087	−6.05603											
Nuclear texture complexity SD (80% based)	4.60160		6.52672	9.33385											

UDH: Usual ductal hyperplasia, ADH: Atypical ductal hyperplasia, LG-DCIS: Low-grade ductal carcinoma *in situ*, IM-DCIS: Intermediate-grade ductal carcinoma *in situ*, IDG: Integrated diffusion gradient, IDG1: Ratio of nucleus area size and rectangle box area (long × short axis), IDG2: Ratio of nucleus 3D volume to cuboid volume, IDG3-8: Total volume over 6 threshold intensity levels, IDG9-14: Increased volume over each threshold intensity level, IDG15-20: Counts for each threshold intensity level cluster, IDG21-26: Image fractal dimensions for each threshold intensity level, GLCM: Gray level co-occurrence matrix, SD: Standard deviation