

# Risk biomarker assessment for breast cancer progression: Replication precision of nuclear morphometry

N. Poulin<sup>a,b,\*</sup>, A. Frost<sup>c</sup>, A. Carraro<sup>a</sup>, E. Mommers<sup>b</sup>, M. Guillaud<sup>a</sup>, P.J. van Diest<sup>b</sup>, W. Grizzle<sup>c</sup> and S. Beenken<sup>c</sup>

<sup>a</sup> *Cancer Imaging, BC Cancer Research Center, Vancouver, Canada*

<sup>b</sup> *Institute for Pathology, Free University of Amsterdam, The Netherlands*

<sup>c</sup> *Departments of Surgery and Pathology, University of Alabama at Birmingham, UK*

**Abstract.** Nuclear morphometry is a method for quantitative measurement of histopathologic changes in the appearance of stained cell nuclei. Numerous studies have indicated that these assessments may provide clinically relevant information related to the degree of progression and malignant potential of breast neoplasia. Nuclear features are derived from computerized analysis of digitized microscope images, and a quantitative Feulgen stain for DNA was used. Features analyzed included: (1) DNA content; (2) nuclear size and shape; and (3) texture features, describing spatial features of chromatin distribution. In this study replicated measurements are described on a series of 54 breast carcinoma specimens of differing pathologic grades. Duplicate measurements were performed using two serial sections, which were processed and analyzed separately. The value of a single feature measurement, the nuclear area profile, was shown to be the strongest indicator of progression. A quantitative nuclear grade was derived and shown to be strongly correlated with not only the pathologic nuclear grade, but also with tubule formation, mitotic grade, and with the overall histopathologic grade. Analysis of replication precision showed that the standard methods of the histopathology laboratory, if practiced in a uniform manner, are sufficient to ensure reproducibility of these assessments. We argue that nuclear morphometry provides a standardized and reproducible framework for quantitative pathologic assessments.

Keywords: Image analysis, image cytometry, reproducibility, standardization

## 1. Introduction

Nuclear morphometry relates to the computerized analysis of digital microscope images of stained cell nuclei, and is used to characterize pathological changes in the appearance of neoplastic cells. The Feulgen stain provides a standardized, quantitative measure of DNA content, as well as a measure of DNA density at each point in the cell nucleus. Shape features describe alterations in the size and shape of cell nuclei. DNA content is measured as the sum of the optical density of stain within the cell nucleus. Texture features describe the distribution of DNA density within the cell nucleus, and reflect changes in chromatin structure associated

with worsening lesion grade (i.e. more poorly differentiated and highly proliferative lesions).

Promising intermediate endpoints have been identified for breast carcinoma progression, and include elements of the Nottingham prognostic index, cytometric DNA content, receptors of the EGF-Erb-B family, estrogen receptors, TGF- $\alpha$ , p53, PCNA, and other proteins involved in control of proliferation (reviewed in [2]). In this context, nuclear morphometry is an attractive biomarker since it is a direct and quantitative measure of an established prognostic marker in breast tissue: the pathologic appearance of the stained cell nucleus. In this report we show that this method provides a quantitative framework for pathologic assessments, and may be useful in the definition of a continuous index of progression for breast carcinoma.

The importance of inter-observer reproducibility in the histopathologic evaluation of breast carcinoma has received considerable attention. Various grading sys-

---

\*Corresponding author: N. Poulin, Kamer OE53, Academisch Ziekenhuis, Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands. Tel.: +31 20 4444025; Fax: +31 20 4442964; E-mail: NM.Poulin@vumc.nl.

tems have been used previously, with variable degrees of reproducibility reported by acknowledged experts in the field. Using WHO criteria, Delides et al. [5] compared grading of 158 tumors by six pathologists at different institutes and found complete agreement in only 23 cases. Using WHO criteria Hopton et al. have reported 78% complete agreement in a study of 874 tumors [11].

Similarly Davis et al. conducted a trial with 1537 patients, with 75 contributing pathologists from local hospitals, and found complete agreement in only 54% of cases [4]. Notwithstanding this wide range of variations, histopathologic grade was determined to have independent prognostic significance when controlled for other clinical parameters.

More recently Elston's modification of the Scarff–Bloom–Richardson grade has gained wide acceptance [7]. The system is referred to as the Nottingham histologic grade, and is used to define a three tiered grading system based on the evaluation of nuclear grade, degree of tubule formation, and mitotic index. The system has been rigorously defined and careful attention has been paid to standardization and reproducibility.

The Nottingham group reports inter-observer agreement on the order of 70–84% [7,16], and shows significant prognostic associations in large cohorts of patients. Similar findings have been reported by other groups using the Nottingham grade [3,9,10]. Interestingly Harvey et al. [10] have also reported that cytometric measurement of DNA content (ploidy) was correlated to grade, and was more discriminating than histologic grade in distinguishing different outcomes. In general the studies tend to report the smallest inter-observer agreement for determination of nuclear grade, and taken together these results indicate that quantitative measures of nuclear grade may provide significant improvements in the prediction of outcome.

Previous work by our group has indicated that a continuous spectrum of nuclear changes occurs over the course of breast carcinogenesis [13]. Aubelle et al. have shown that measures of quantitative nuclear grade may provide improved prognostication in node positive breast carcinoma [1]. Hoque et al. [12] have shown that nuclear morphometry may be useful in the prediction of recurrence of DCIS after surgery and radiation. Fabian et al. have shown that these measures may be used to identify cohorts of patients at very high short term risk for the development of breast cancer [8]. Similarly Mommers et al. have shown that analysis of morphometric features in breast hyperplasia may provide significant improvements in risk assessment for development of carcinoma [14].

These and other studies have shown the value of a quantitative imaging framework for assessment of cytologic atypia in breast carcinoma. However concerns have been raised with the use of thin tissue sections, related to the reproducibility and standardization of these measures. In this report we describe a validation study of these measurements on thin tissue sections, processed according to the standard procedures of the institutional histopathology laboratories. The study was performed on a series of replicated specimens which were processed independently, and which reflected the spectrum of changes seen in different grades of breast carcinoma.

## 2. Materials and methods

### 2.1. Specimens

Formalin fixed, paraffin-embedded specimens were analyzed in a series of 54 breast cancer specimens. Elston's modification of the Scharff–Bloom–Richardson system was used for histopathologic classification of these cases into nine categories. The system is referred to as the Nottingham grade [16], and relies on analysis of three histopathologic features: (1) nuclear grade, (2) degree of tubule formation, and (3) mitotic index. Lesions are given a score between 1 and 3 in each category, with 3 representing the score with the worst prognostic association (i.e. higher nuclear grade, less tubule formation, and higher mitotic index). The Nottingham grade is a sum of these scores, and the breakdown of cases with respect to this classification is summarized in Tables 1 and 2.

Thin tissue sections were cut at a thickness of 4–5 micrometers. A direct measure and calibration of section thickness was not considered practical, and standard procedures of the institutional histopathology laboratories were used. Paraffin sections were melted on glass slides at 60°C for 30 minutes, sections were deparaffinized in two changes of xylene, rehydrated through graded ethanol, and stained according to a modified Feulgen procedure.

Table 1

Breakdown of the number of cases with respect to components of the Nottingham grade

Grade	Nuclear	Mitotic	Tubule formation
1	15	22	14
2	14	9	8
3	25	23	32

Table 2

Breakdown of the number of cases in each category of the Nottingham scoring system. Scores from 3–5 are Nottingham grade I, 6–7 are grade II, and 8–9 are grade III

Nottingham score	Cases	Nottingham grade	Cases
3	12	I	18
4	3		
5	3		
6	5	II	9
7	4		
8	14	III	27
9	13		

## 2.2. Staining

The staining solution was prepared:

- (1) By boiling 0.25 g thionin (Sigma chemical company) in 220 ml distilled water for 5 minutes, and cooling until lukewarm.
- (2) The following ingredients were added: 220 ml tertiary butanol, 65 ml 1 N HCl, and 4.325 g sodium bisulfite.
- (3) The solution was stirred for 1.5 hours, and left at room temperature for at least 20 hours to equilibrate.

The staining procedure was performed under temperature controlled conditions, at 20°C for steps (1)–(5) below:

- (1) 10-minute post fixation in Boehm–Sprenger fixative (320 ml methanol, 60 ml formaldehyde, 20 ml acetic acid).
- (2) 60-minute hydrolysis of specimen in 5 N HCl.
- (3) Immersion in the stain solution for 1 hour.
- (4) Staining was differentiated for specificity for the cell nucleus, by rinsing in a bisulfite solution (7.5 g sodium bisulfite, 1425 ml distilled water, 75 ml 1 N HCl) for 10 minutes.
- (5) Stained specimens were rinsed in two changes of distilled water, dehydrated, and mounted with coverslips using CytoSeal (ProSciTech, Australia).

Regions of interest on the slide were identified by the study pathologists (A.F., P.v.D.). These included areas of uninvolved epithelium (normal), atypical hyperplasia (ADH), ductal carcinoma *in situ* (DCIS), and invasive ductal carcinoma (IDC). Images of at least 200 cells were collected in each category where possible. All regions of interest on the slide were sampled in a uniform manner. Isolated and non-overlapping cell

images were selected by the operator at convenience (i.e. where available), subject to the criterion that the slide was sampled uniformly over indicated regions of interest.

## 2.3. Morphometry

Cell feature measurements were performed as previously described [6,14], involving the interactive selection of digital nuclear images, identification of internal diploid control cells (lymphocytes), and the calculation of normalized texture features. The resolution of nuclear images was 0.34 micrometers (pixel spacing at the specimen plane).

89/128 feature measurements performed by the Cyto-Savant device (CCABC) were considered informative. Uninformative features relate to a number of quality control parameters, to features which are not used due to analytical concerns, and to the stage and screen coordinates of the cell. Statistical analyses of cell features were performed using the mean values of each feature measurement for each case. Separate analyses were also performed using the coefficients of variation (standard deviation/mean) as parameters for each case.

## 2.4. Analysis

Replication precision was evaluated using Wilcoxon matched pairs analysis, and independently using ANOVA with repeated measures design. Correlations between replicated measurements were analyzed, and the Pearson correlation statistic  $R^2$  was used to determine the overall proportion of the sample variance which was due to replication error ( $1 - R^2$ ).

ANOVA was used with pathologic nuclear grade as the independent (grouping) variable, in order to determine a list of features which showed significant effects. Principle component analysis was performed on variables from this list, for the purpose of dimensionality reduction – principle components were evaluated separately on standardized variables for means and c.v.'s of significant features. Stepwise multivariate linear regression was performed using the un-rotated principle components, and using pathologic nuclear grade as the dependent variable. Normal cells were assigned a nuclear grade of 0 in these analyses.

### 3. Results

#### 3.1. ANOVA

For measurements performed on IDC lesions, analysis of variance was used to determine a list of variables which showed significant univariate effects in the analysis of 4 categories: normal (uninvolved epithelium), and IDC of low (1), intermediate (2) and high (3) nuclear grades. 68/89 informative variables were identified with  $F$  values greater than 5.7. These corresponded to variables with significant effects with  $p < 0.001$ .

Features related to the mean nuclear size (area, radius, maximum radius) had the largest effects, with  $F = 146, p < 10^{-6}$  for mean nuclear area. Mean DNA index showed a significant effect but was less informative, with  $F = 38, p < 10^{-6}$ . This is an expected result

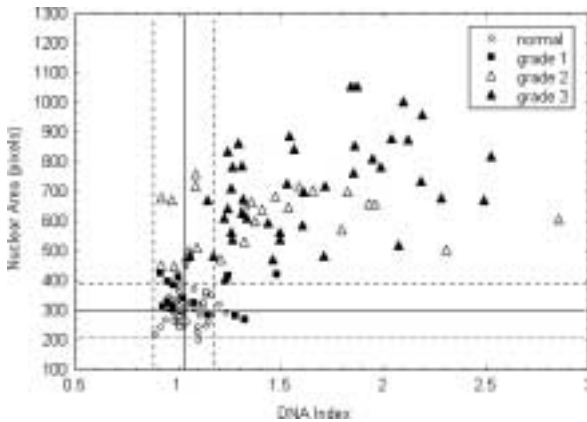


Fig. 1. Mean nuclear area vs mean DNA index in breast carcinoma specimens. The open circles in the diploid range represent normal epithelial cell measurements combined from all biopsies. 95% confidence bands (dashed lines) for the estimation of the means (solid line) are shown for normal cells (open circles).

due to the fact that in thin histologic sections the DNA content measurement is compromised, since nuclei are not intact but are in fact sectioned to varying degrees.

A plot of nuclear area vs DNA index is shown in Fig. 1. Normal cell measurements in the diploid range are shown as open circles. The mean (solid line) and 95% confidence intervals (dashed lines) are shown for measurement of normal cells.

As evident in the figure, many cases with grade 1 nuclei are not distinguished from normal cell measurements using these variables. Grade 2 and grade 3 categories both show a broad range over these measurements, and are not well separated with these parameters. However they are clearly separated from normal and grade 1 nuclei.

Post hoc comparisons of feature means were examined for each variable, and are summarized in Table 3 for selected variables which showed significant univariate effects. Significance tests listed in the table are for the distinction between overlapping groups:

- (1) normal vs grade 1,
- (2) grade 1 vs grade 2,
- (3) grade 2 vs grade 3.

All variables with significant effects were significantly correlated with nuclear area ( $p < 10^{-6}$ ).

A similar analysis is summarized in Table 4, where ANOVA was performed using the c.v.'s of feature measurements as input parameters. 47/89 informative parameters were determined to have significant effects with  $p < 0.001$ .

#### 3.2. Factor analysis

Factor analysis of variables with significant effects ( $p < 0.001$ ) was performed. This is a method for reduction of the number of variables in the analy-

Table 3

Post hoc comparison of group means. Tukey's honest significant differences (HSD) for unequal  $N$  were used to determine significance values for differences between the indicated overlapping groups. The  $F$ -value is listed for the distributions of each feature mean, and significant correlations with mean nuclear area are also shown

	Normal vs grade 1 ( $p$ value)	Grade 1 vs grade 2 ( $p$ value)	Grade 2 vs grade 3 ( $p$ value)	$F$ value (ANOVA)	Correlation with area
Nuclear area	0.49	0.000008	0.0017	146	1.0000
DNA index	0.97	0.006	0.11	38	0.687
Range extreme	0.0868	0.00004	0.67	80	0.585
Fractal area	0.15	0.00001	0.0001	129	0.984
Run length	0.06	0.07	0.4	46	0.963
Contrast	0.93	0.000008	0.99	52	0.811

Table 4

*Post hoc* comparisons of coefficients of variation. Tukey's honest significant differences (HSD) for unequal  $N$  was used to determine significance values for differences between the indicated groups. The  $F$ -statistic and correlations with mean area are also noted

	Normal vs grade 1 ( $p$ value)	Grade 1 vs grade 2 ( $p$ value)	Grade 2 vs grade 3 ( $p$ value)	$F$ value (ANOVA)	Correlation with mean area
Nuclear area	0.32	0.58	0.22	23	0.514
DNA index	0.15	0.05	0.03	54	0.651
medDNAcomp	0.06	0.01	0.98	40	0.611
Fractal dim	0.13	0.0003	0.56	68	0.818
Run length	0.06	0.07	0.4	46	0.780
densdarkspot	0.006	0.005	0.27	80	0.471

Table 5

Summary of factor analyses. Tukey's honest significant differences (HSD) for unequal  $N$ . Significance values for *post hoc* comparison of factors are shown for overlapping groups in the last three columns

	Variance explained (%)	$F$ value (ANOVA)	Correlation with area	Normal vs grade 1 ( $p$ value)	Grade 1 vs grade 2 ( $p$ value)	Grade 2 vs grade 3 ( $p$ value)
F1(mn)	49	109	-0.93	0.12	$10^{-6}$	0.17
F2(mn)	12	5	-0.32	0.39	0.04	0.99
F3(mn)	10	2	0.06 (n.s.)	0.24	0.66	0.99
F4(mn)	0.6	4	0.10 (n.s.)	0.92	0.81	0.05
F1(cv)	37	99	0.82	0.006	0.0002	0.31
F2(cv)	14	1	0.24	0.96	0.90	0.23
F3(cv)	7	2	0.003 (n.s.)	0.7	0.05	1.0
F4(cv)	6	4	0.06 (n.s.)	0.04	0.40	0.80
QNG	80	218	0.95	0.005	$10^{-6}$	0.0003

sis, yielding a transformation of input variables to a smaller set, which better summarizes the variance in the dataset.

Factors are constructed by replacement of the variables with linear combinations of the original variables. The new variables are chosen sequentially, such that they are independent (not correlated), and such that they represent the direction which maximizes the remaining variance of the data. This method accounts for the contributions from highly correlated variables since these are grouped in linear combinations to form the new factor. Unrotated principle components were derived using standardized variables for these analyses.

Separate analyses were performed using two sets of input variables, the means and coefficients of variation for each case. These analyses are summarized in Table 5, which shows the *post hoc* tests of differences in nuclear grades for each variable. In Table 5 factor 1 from the analysis of feature means is denoted: F1(mn), and the first factor from the analysis of coefficients of variation is denoted F1(cv).

### 3.3. Regression

A quantitative nuclear grade (QNG) was finally assigned through multiple regression using all factors, with nuclear grade as the independent variable. Regression analysis with stepwise variable selection yielded a model for quantitative nuclear grade, including the first two factors from analysis of feature means and c.v.'s, with a multiple  $R^2 = 0.80$ . Independent significance was obtained only for the first factor from analysis of feature means. The value of  $1 - R^2$  represents the fraction of variance which is due to departures from the regression line (residuals).

Figure 2 illustrates the correlation between the QNG and nuclear area, and compares the separability of nuclear grades with respect to these parameters. There are a number of observations regarding the distribution of the quantitative nuclear grade:

- there is high degree of overlap between pathologic nuclear grades,

- (b) the parameter QNG is heterogeneously distributed across cases with grade 2 and grade 3 nuclei,
- (c) there is clearly a class of grade 1 nuclei which are indistinguishable from normal cells using this grade.

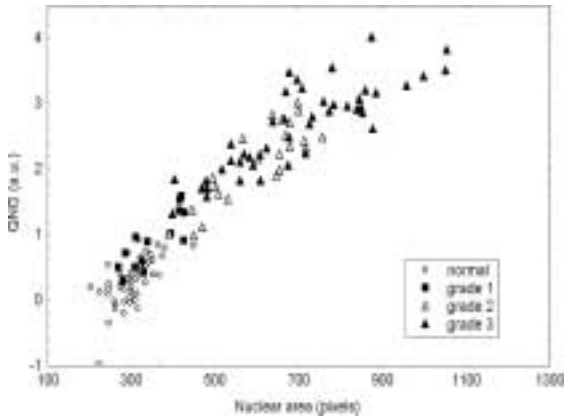


Fig. 2. Correlation of quantitative nuclear grade (QNG) with mean nuclear area.

The apparent overlap of these populations was investigated by inspection of extreme cases of discrepancy between QNG and pathologic nuclear grade. Recall that the plotted parameters are derived from the *means* and *c.v.'s* of feature measurements, calculated over the population of tumor cells for each slide.

In Fig. 3 a histogram of area measurements is shown for a populations of cells from a single case. In both the DCIS and invasive components, smaller subpopulations of higher grade nuclei are apparent. The use of the values of the slide mean and *c.v.'s* for each parameter tends to obscure the presence of more rare high grade cells, whereas even small foci of these cells are diagnostic for the determination of high nuclear grade.

### 3.4. Replication precision

There are known limitations to morphometric measurements using thin tissue sections, most notably due to the fact that nuclei are not intact but are themselves sectioned. Variations in section thickness between specimens, and within the same specimen, may

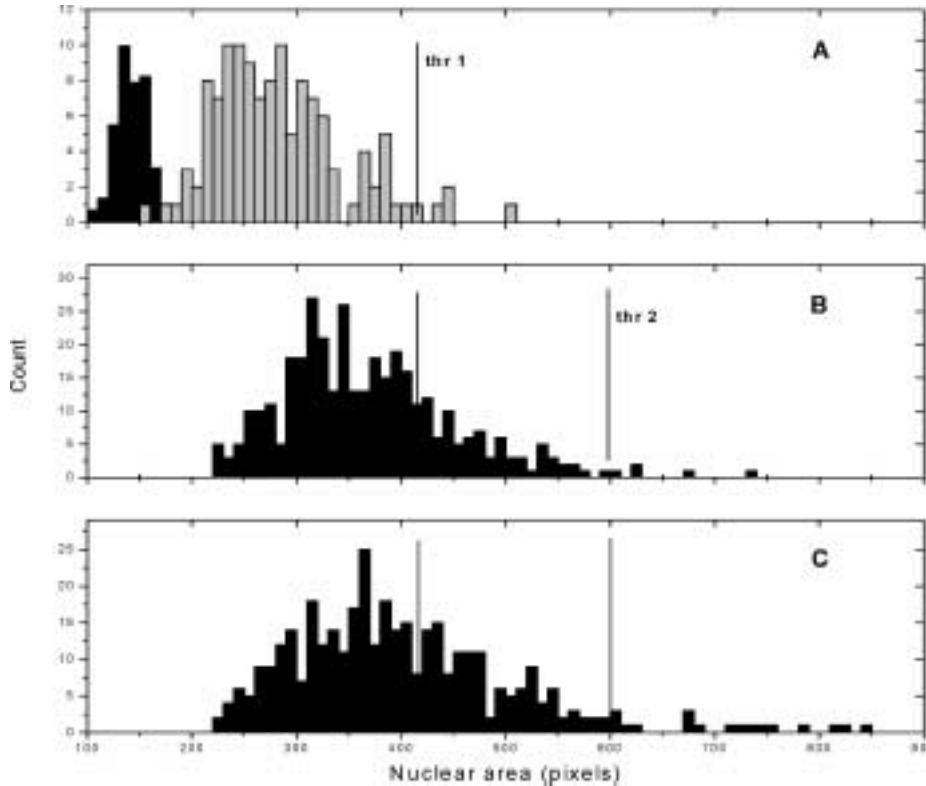


Fig. 3. Histogram of nuclear area for individual cell nuclei within a single case. A: lymphocyte (solid) and normal (uninvolved) cells (grey). B: DCIS component of the lesion. C: Invasive component. The vertical line at thr1 represents the cutoff value for grade 1 nuclei (mean + std. error), and at thr2 the cutoff value for grade 2 nuclei.

be the largest source of error in cell feature measurements. For this reason the study was designed to investigate reproducibility: two serial sections were processed and measured independently for the majority of specimens in this series.

Using Wilcoxon matched pairs tests, and separate ANOVA tests with repeated measures design, no significant systematic difference between replicated measurements was found for any feature mean ( $p > 0.05$ ). Similarly these tests were not significant for differences in values of the quantitative nuclear grade (QNG). For feature c.v.'s, 3/47 informative parameters showed significantly lower values in the second replicate – however the significance of these differences was marginal ( $0.01 < p < 0.05$ ).

The precision of the QNG may be determined through analysis of correlations between replicated measurements. The correlation coefficient between the two replicates of QNG,  $R$ , is 0.967. The value of the expression:  $(1 - R^2)$  represents the fraction of the total variance which is not due to correlation, and may be regarded as a measure of the average replication error. For the QNG, replication error was 6%, and for mean nuclear area, the replication error was 9%.

The most serious problem with replication precision is related to the sampling issues noted in Fig. 3 above, which showed that small subpopulations of high grade cells determine the pathologic nuclear grade in many of these lesions. The replication precision of the sampling strategy was investigated by analyzing the frequency of cell subpopulations defined by thresholds on nuclear area, as shown in Fig. 3 as *thr1* and *thr2*.

Cells with nuclear area below the first threshold (*thr1*) were defined as low grade, cells with area between *thr1* and *thr2* were defined as medium grade, and cells larger than *thr2* were considered high grade. The frequency of low, medium, and high grade cells was calculated for each slide, and tested for reproducibility by analysis of correlations between replicated measurements. In this manner the replication error ( $1 - R^2$ ) was determined to be 34%, 44% and 24% for the frequency of low, medium, and high grade cells, respectively.

In Fig. 4 the replication error of the QNG is categorized with respect to differences observed between normal and abnormal cells. In the cases with the greatest difference between replicates, the error is comparable with the range of variability across normal specimens. This range is indicated in the figure by the dashed lines outlining the region: mean  $\pm 1$  standard deviation, for observations of normal cells).

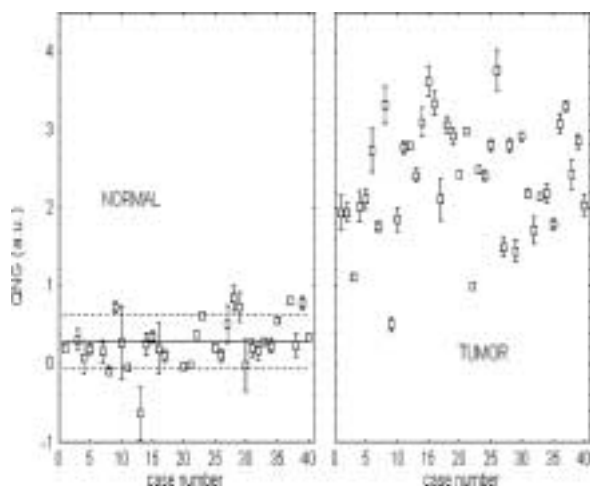


Fig. 4. Reproducibility of quantitative nuclear grade. Minimum and maximum values of the QNG are indicated by the horizontal lines for each case. Variability of measurements of normal cells in this series is indicated by the dashed lines on either side of the mean (solid line), representing  $\pm 1$  standard deviation from the mean.

### 3.5. Quantitative nuclear grade

In Fig. 5 the distribution of the quantitative nuclear grade is shown with respect to different components of the Nottingham scoring system. Interestingly, this cytologic grading system, developed with reference to the nuclear grade, reflects changes in the other parameters as well. In particular the degree of tubule formation is correlated with the quantitative nuclear grade. Cells in the highest mitotic index category are apparently of lower nuclear grade than cells with intermediate mitotic index.

In Fig. 6 it is apparent that the overall histopathologic grade of the lesion is also strongly correlated to the quantitative nuclear grade. In the figure the distribution of the QNG is compared to that of the mean nuclear area. Given the degree of overlap it is evident that QNG defines essentially three categories of histopathologic grade: low, intermediate, and high. QNG yields greater distinction than mean nuclear area between normal and grade 1 nuclei. It may also be significant that lesions with a Nottingham score of seven are closest to the highest grade category using this system.

## 4. Discussion

A straightforward algorithm was used to define a quantitative nuclear grade for breast carcinoma specimens, using cytometric feature means and coefficients

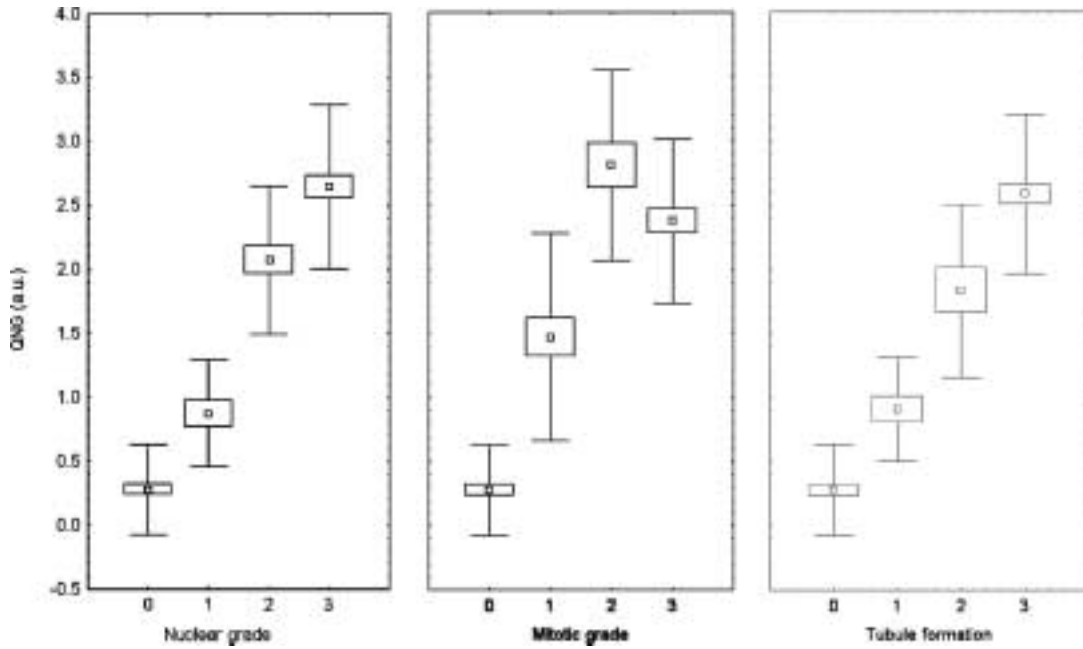


Fig. 5. Distribution of quantitative nuclear grade with respect to components of the Nottingham scoring system: nuclear grade, mitotic index, and degree of tubule formation. The boxes associated with each point represent the region  $\pm 1$  standard error from the mean, and the horizontal lines (whiskers) represent the region  $\pm 1$  standard deviation.

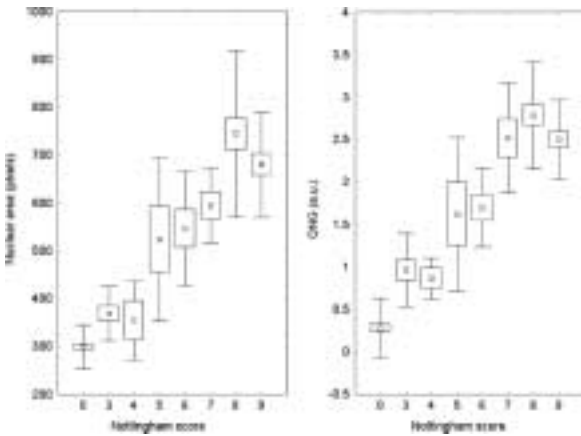


Fig. 6. Distribution of mean nuclear area and quantitative nuclear grade, with respect to categories of the Nottingham scoring system.

of variation as input parameters. ANOVA was used to identify a list of variables with significant effects with respect to pathologic nuclear grade. Principle component analysis was used for reduction of dimensionality, and stepwise multiple linear regression on pathologic nuclear grade was used to define a function of these principle components, the quantitative nuclear grade.

The selection of individual features for inclusion in principle component analysis was based on a threshold

value of significance: i.e.  $p < 0.001$ . Principle components analysis showed that the features were not independently distributed, but were in fact highly intercorrelated. For this reason corrections for multiple univariate tests were not considered necessary, since at most four independent components were identified as the major contributors to variance within the data. The data are not considered to be “over-fitted”, since no attempt is made to develop classifiers in this analysis.

The distribution of the parameter QNG over the individual components which comprise the Nottingham grade is shown in Fig. 5. This parameter discriminates between different pathologic nuclear grades, but it is interesting that this score is also correlated with degree of tubule formation and mitotic index grades. Moreover quantitative nuclear grade is also correlated to Nottingham score, and may represent a continuous index of progression.

The overlap between the categories of nuclear grade in this study has been investigated in extreme cases. In particular there are a number of grade 3 lesions, which appear to fall within the diploid, low grade range of these measurements. Inspection of these cases showed that small subpopulations of higher grade nuclei have provoked the diagnosis of nuclear grade 3. The measurement of average feature values, over pre-



dominantly lower grade populations, does not represent these small subpopulations. It is likely that this intra-lesion heterogeneity accounts for a great deal of the apparent overlap between nuclear grade categories.

It is important to note that measurements on thin tissue sections have known limitations. Overlapping cells are frequent and cannot be adequately measured. Variations in section thickness arise and have been estimated to be on the order of 1 micron. This may skew area estimations by as much as 10%, as seen in this study for normal cells, and as reported in the literature [6]. Aneuploidy may not be accurately assessed in thin tissue sections, since cells are not intact but are sectioned in random cross section.

The uniform sampling strategy employed in this study has poor reproducibility particularly for rare subpopulations of cells (e.g., 24% error for high grade cells in this series), and may be due in part to errors in area estimations, lack of measurable cells, and lack of an explicitly automated sampling procedure. While systematic random sampling is the preferred method for representation of cell populations, with interactive methods this is very time consuming, and rare events may still be inadequately sampled. An adaptive sampling strategy, which focuses attention to the areas of highest nuclear grade, may prove to be optimal. Future validation studies should take explicit account of this problem in terms of instrument design and quality assurance protocols.

Various methods have been discussed in relation to the correction of nuclear area and DNA content measurements in thin tissue sections. A survey of the literature shows that the application of these techniques is variable and subject to numerous restrictive assumptions. In this study no such corrections were attempted for area and DNA content measurements. Nuclei which were severely sectioned, e.g., with diameter smaller than 5 micrometers, were excluded from analysis, but sectioning artifact remains the most serious analytical problem to QNG estimations on thin tissue sections.

Nevertheless, replication precision measurements on the specimens in this series show that the measurements are accurate within an acceptable margin of replication error (e.g., 6% for QNG), and that histopathologic grading schemes can be represented within an objective framework. Further analyses of high grade subpopulations are expected to reveal these distinctions with much greater resolution. The study suggests that clinically relevant distinctions can be made with use of the current method.

## Acknowledgements

This work was supported by a grant from the NIH (USA): # N01-CN-35577-03 and by a grant from the Pasman Stichting Foundation (Netherlands).

## References

- [1] M. Aubele, G. Auer, A. Voss, U. Falkmer, L. Rutquist and H. Hoffer, Disease-free survival of node-positive breast cancer patients. Improved prognostication by cytometrical parameters, *Pathology, Research & Practice* **191**(10) (1995), 982–990.
- [2] N.J. Bundred, Prognostic and predictive factors in breast cancer, *Cancer Treatment Reviews* **27**(3) (2001), 137–142.
- [3] L.W. Dalton, D.L. Page and W.D. Dupont, Histologic grading of breast carcinoma. A reproducibility study, *Cancer* **73**(11) (1994), 2765–2770.
- [4] B.W. Davis, R.D. Gelber, A. Goldhirsch, W.H. Hartmann, G.W. Locher, R. Reed, R. Golouh, J. Save-Soderbergh, L. Holloway, I. Russell et al., Prognostic significance of tumor grade in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis, *Cancer* **58**(12) (1986), 2662–2670.
- [5] G.S. Delides, G. Garas, G. Georgouli, D. Jiortziotis, J. Lecca, T. Liva and J. Elemenoglou, Intralaboratory variations in the grading of breast carcinoma, *Arch. Pathol. Lab. Med.* **106**(3) (1982), 126–128.
- [6] A. Doudkine, C. MacAulay, N. Poulin and B. Palcic, Nuclear texture measurements in image cytometry, *Pathologica* **87**(3) (1995), 286–299.
- [7] C.W. Elston and I.O. Ellis, Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer, *Histopathology* **19** (1991), 403–410.
- [8] C.J. Fabian, B.F. Kimler, C.M. Zalles, J.R. Klemp, S. Kamel, S. Zeiger and M.S. Mayo, Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model, *Journal of the National Cancer Institute* **92**(15) (2000), 1217–1227.
- [9] H.F. Frierson, R.A. Wolber, K.W. Berean, D.W. Franquemont, M.J. Gaffey, J.C. Boyd and D.C. Wilbur, Interobserver Reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma.
- [10] J.M. Harvey, N.H. de Klerk and G.F. Sterrett, Histologic grading in breast cancer: Interobserver agreement, and relation to ploidy and other prognostic parameters, *Pathology* **24** (1992), 63–68.
- [11] D.S. Hopton, J. Thorogood, A.D. Clayden and D. MacKinnon, Observer variation in histological grading of breast cancer, *Eur. J. Surg. Oncol.* **15**(1) (1989), 21–23.
- [12] A. Hoque, S.M. Lippman, I.V. Boiko, E.N. Atkinson, N. Sneige, A. Sahin, D.M. Weber, S. Risin, M.D. Lagios, R. Schwarting, W.J. Colburn, K. Dhingra, M. Follen, G.J. Kelloff, C.W. Boone and W.N. Hittelman, Quantitative nuclear morphometry by image analysis for prediction of recurrence of ductal carcinoma in situ of the breast, *Cancer Epidemiology, Biomarkers & Prevention* **10**(3) (2001), 249–259.

- [13] E.C. Mommers, N. Poulin, J. Sangulin, C.J. Meijer, J.P. Baak and P.J. van Diest, Nuclear cytometric changes in breast carcinogenesis, *Journal of Pathology* **193**(1) (2001), 33–39.
- [14] E.C.M. Mommers, D.L. Page, W.D. Dupont, P. Schuyler, A.M. Leonhart, J.P.A. Baak, C.J.L.M. Meijer and P.J. van Diest, Prognostic value of morphometry in patients with normal breast tissue or usual ductal hyperplasia of the breast, *Int. J. Cancer* **95** (2001), 282–285.
- [15] J.P. Rigaut and A. Persoz, The “corpuscle” sterological problem – re-evaluation using slab fragment volumes and application to the correction of DNA content histograms from sections of spherical nuclei, *J. Microsc.* **156** (1989), 371–382.
- [16] P. Robbins, S. Pinder, N. de Klerk, H. Dawkins, J. Harvey, G. Sterrett, I. Ellis and C. Elston, Histologic grading of breast carcinoma. A study of interobserver agreement, *Hum. Path.* **26**(8) (1995), 873–879.