Self-Affinity and Lacunarity of Chromatin Texture in Benign and Malignant Breast Epithelial Cell Nuclei

Andrew J. Einstein,[1,2,*] Hai-Shan Wu,[2] and Joan Gil[2]
[1]*Department of Biomathematical Sciences, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1023, New York, New York 10029*
[2]*Department of Pathology, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1194, New York, New York 10029*
(Received 11 September 1996)

Methods are presented for characterizing the self-affinity and lacunarity of arbitrarily shaped images. Chromatin appearance in breast epithelial cell nuclei is shown to be statistically self-affine. Spectral and Minkowski dimensions are lesser in nuclei of malignant cases than in nuclei of benign cases, and lacunarity further quantifies morphologic differences such as chromatin clumping and nucleoli. Fractal texture features are used as the basis for an accurate cytologic diagnosis of breast cancer. [S0031-9007(97)04986-7]

PACS numbers: 87.10.+e, 87.56.Fc

Contemporary standards in oncology typically require a pathologic diagnosis prior to instituting therapy, and, thus, the definitive role in cancer diagnosis is borne by the pathologist. Despite the introduction of promising molecular biomarkers, diagnosis is still made by the pathologist largely on the basis of a subjective morphologic assessment of physical properties of cells and nuclei such as size, shape, adhesion, regularity, and chromatin appearance. Factors considered in evaluating chromatin include whether it takes a fine or coarse appearance, if it is clumped, whether marginated chromatin and/or nuclear voids are present, and the size and number of nucleoli. While, in most cases, diagnosis is straightforward for an experienced pathologist, in some cases, it is not. The diagnosis of breast cancer poses a particular challenge. Fine-needle aspiration (FNA) cytology, which involves inserting a needle into the breast and aspirating individual cells from suspicious tissue, has become an increasingly important diagnostic procedure in the management of breast masses. An accurate diagnosis using FNA is highly dependent on how experienced the aspirator and cytopathologist are, and high false negative rates have been reported [1].

Scale-invariant behavior is characteristic of many biological systems, ranging from microvascular network formation to heart interbeat intervals [2]. Recent studies have shown that aspects of morphology considered by pathologists, such as the shapes of cell membranes [3], nuclear membranes [4], and tissues [5], exhibit fractal properties. Here, we investigate scale-invariant properties of chromatin appearance in microscopic images of breast epithelial cell nuclei obtained by cytology, and show how they can be used as the basis for a more objective diagnosis between benign and malignant breast epithelial cell lesions. Chromatin appearance in nuclear images is manifested in terms of texture, the spatial distribution of grey values. Several studies have attempted to quantify texture in cytologic specimens, for both diagnostic [6] and prognostic [7] purposes.

As is illustrated in Fig. 1, a nuclear image can be viewed as a surface for which the x and y coordinates represent position and the z coordinate represents grey level. We evaluate the self-affinity of these surfaces, and characterize them by fractal dimensions and lacunarity. If nuclear images can be shown to be (statistically) self-affine, then these fractal parameters are an appealing approach to texture characterization. Pentland [8] has shown a high correlation between human visual roughness perception and fractal dimension, and, thus, fractal dimension is a good candidate for quantitating the nuclear irregularity subjectively observed by pathologists. Moreover, lacunarity [9] measures the largeness of gaps or holes. Since nucleoli and chromatin voids form large "holes" in cytologic images, and increases in nucleolar size and irregularity as well as large voids tend to be indicative of malignancy [10], lacunarity should also be of diagnostic utility.

We obtained cytology specimens from 41 patients. The specimens were stained using the ultrafast Papanicolaou protocol [11]. Twenty-two patients were diagnosed with invasive ductal carcinoma and 19 as benign; each

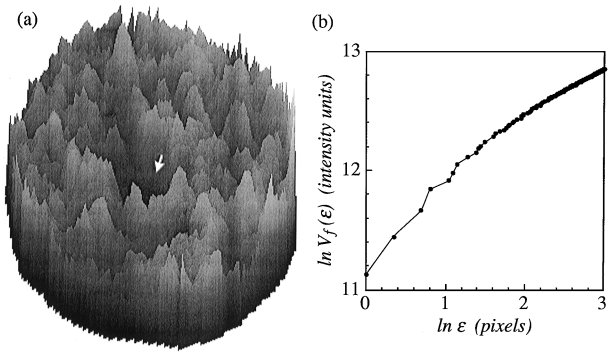

FIG. 1. (a) Surface plot of a malignant breast epithelial cell nucleus; note the nucleolar "pit," marked by an arrow. Fractal dimensions, defined in the text, are $D_{\mathrm{MB}} = 2.53$ (as determined by least-squares fit of all 145 values of ε) and $D_S = 2.85$. (b) The associated plot of $\ln V_f(\varepsilon)$ versus $\ln \varepsilon$. $r^2 = 0.97$, indicating that chromatin texture is self-affine.

diagnosis was made by a cytologist and independently confirmed on the basis of surgical biopsy findings. Specimens were viewed on an image analysis system under oil-immersion light microscopy, and an average of 64 randomly selected nuclear images were segmented for each patient by an individual blinded as to the cytologic and histopathologic diagnoses. Since, in cytologic specimens, cells lie compressed between the slide and the cover slip, these images provide planar representations of complete nuclei. Images were normalized [12] to compensate for possible differences in staining and lighting conditions.

Several fractal dimensions can be used to characterize self-affine surfaces, each one contributing to an overall description [13]. We characterized the nuclear images in the spatial domain with a Minkowski-Bouligand dimension D_{MB}, and in the frequency domain with a spectral exponent β. Minkowski dimensions were determined using a modification of the variation method of Dubuc *et al.* [14]. For a surface $z = f(x, y)$ at a given (x', y'), the ε oscillation is defined as the difference between the extreme values of f in a neighborhood of (x', y'), i.e.,

$$V_\varepsilon(x', y') = \max f(x, y) - \min f(x, y) .$$

We determine the extrema over all (x, y) on the surface such that $(x - x')^2 + (y - y')^2 \leq \varepsilon^2$. The ε variation $V_f(\varepsilon)$ of f is the sum of $V_\varepsilon(x', y')$ over the entire surface. The variation method estimates D_{MB} by three minus the slope of the least-squares line fitting the plot $[\ln \varepsilon, \ln V_f(\varepsilon)]$ over a suitable range of ε. For the nuclear images, we determined $V_f(\varepsilon)$ at 145 values of ε, ranging between 0.134 μm (1 pixel) and 2.69 μm (20 pixels).

The plots $[\ln \varepsilon, \ln V_f(\varepsilon)]$, an example of which is shown in Fig. 1, can be used for two purposes: to assess the validity of the self-affine model, and to estimate D_{MB}. Values of r^2, a standard measure of goodness of fit, ranged from 0.836 to 0.996 for the 2621 nuclei, with a median of 0.962. The good linear fit implies that chromatin appearance is indeed self-affine over the range of resolutions considered. That nuclear appearance is fractal is highly suggestive that the three dimensional organization of nuclear chromatin is also fractal. Supportive of this, Pentland [8] has shown that under certain assumptions, a three-dimensional surface is fractal if and only if its image intensity surface is fractal. Thus, these results extend the scope over which genetic material may have fractal organization, from single chromosomes [15] to nuclear organization.

For an individual patient, we can compute the mean value of D_{MB} over all of the nuclei segmented from that patient, which we refer to as the patient mean of D_{MB}. The average of the individual patient means for a whole group of patients represents an average fractal chromatin texture property of that group, and is referred to as the overall mean for the group. The overall mean D_{MB} was 2.527 for the benign cases and 2.510 for the malignancies. The statistical significance of this observed difference between the overall mean D_{MB} of the benign and malignant cases is analyzed using two-way analysis of variance (ANOVA), which here yields a significance level p of 0.067, corresponding to borderline significance. D_{MB} estimated from a smaller range of resolutions (21 values of ε between 0.134 and 0.940 μm) exhibited a statistically significant difference between benign and malignant cases ($p = 0.044$).

Despite the high coefficients of correlation between $\ln \varepsilon$ and $V_f(\varepsilon)$, a slight concavity can be observed in the $[\ln \varepsilon, \ln V_f(\varepsilon)]$ plots for most nuclei. This concavity may be due to edge effects, or to a slight intrinsic deviation of chromatin texture from mathematical fractality. Consistent with this, we observed that the slopes of the plots, and consequently their associated D_{MB}, depend on the range of ε included in the regression, notwithstanding the high correlation coefficients common to various ranges. D_{MB} were lesser over the smaller range of resolutions (21 ε), with mean dimensions per patient of 2.333 for the benign cases and 2.319 for the malignancies, than over the longer range (145 ε). Rigaut claims that this concavity is present in almost all naturally occurring fractals, and cites several empirical models which have been proposed to incorporate the concavity [16]. Alternatively, some (e.g., Mandelbrot, cited in [17]) prefer to account for concavity with successive values of fractal dimension. Recent studies [18] showing that random processes can generate fractal structures whose log-log plots exhibit curvature provide a mechanism to account for concave log-log plots; however, it is unclear what role, if any, such random processes play in determining chromatin morphology, and how these processes differ between benign and malignant nuclei.

Fractal images are typified by a power spectrum in which there is a $1/f^\beta$ dependence on frequency. Voss [19] has shown that for statistically self-affine fractal Brownian motion, the spectral exponent β is related to a fractal similarity dimension D_S by

$$D_S = (7 - \beta)/2 . \tag{1}$$

Since power spectra are determined from an image's two-dimensional Fourier transform, this approach is limited to rectangular images. Because nuclei have irregular shapes, we developed a new method to determine the fractal dimension of images of arbitrary shape [20]. A nuclear image R_n is embedded in a rectangular array R. The region of R surrounding the actual nuclear image is referred to as R_b; the goal of the iterative procedure is to fill in this background region in such a way that it is similar to the nucleus in the Fourier domain. R_b is initialized such that the statistical properties of the image in R_n are extended to R_b. Following this initialization, the two-dimensional Fourier transform of R is taken, from which its power spectrum is determined. β is estimated by fitting this computed power spectrum to the assumed model in which power spectrum equals $Cf^{-\beta}$, where C is a proportionality constant. The fitted power spectrum is combined with the actual phase of the Fourier transform of R to yield a new frequency domain signal. The inverse Fourier transform of this signal is taken to produce

another spatial domain image R'. The background area R_b' in R' has been determined both from R_b and from spectral properties of the nucleus, and in this sense should be an improvement over R_b. Thus, the image used in the next iteration contains the original nuclear image R_n padded with the improved background signal R_b'. This process is iterated until the change in the coefficients of the $1/f^\beta$ model is sufficiently small, and the final β is used to determine the fractal dimension from (1). In this study, iteration was stopped when the change in β was less than 0.01, or after 200 iterations. We have tested this method with synthetic "nuclei" of known fractal dimension generated using Voss's inverse Fourier filtering method [19], and have found that β is estimated very accurately, e.g., within 0.2% for a nucleus with $D_S = 2.75$.

The overall mean spectral dimension was 2.853 for the benign cases and 2.796 for the malignancies, a difference which was determined to be very highly significant (ANOVA significance level $p = 0.000\,011$). These means exclude 239 of the 2621 nuclei for which the iterative method was considered to have converged incorrectly ($D_S < 2$, i.e., spectral density approaching white noise, or $C > 2000$). Spectral and Minkowski dimensions for individual nuclei were weakly correlated ($r^2 = 0.027$ for the 2382 nonexcluded nuclei), as were patient means of D_S and D_{MB} ($r^2 = 0.018$), indicating that these two dimensions reflect different facets of the self-affine nature of chromatin in breast epithelial cell nuclei. The lower fractal dimensions exhibited in malignancy—in both the frequency and spatial domains—are consistent with the hypothesis advanced by Goldberger and colleagues [21] that disease is associated with a loss of biologic complexity.

Two objects can exhibit different texture but still have the same fractal dimension. As is illustrated in Fig. 2, this difference can often be characterized in terms of lacunarity. We measure weighted lacunarity on thresholded nuclear images using an extension of the gliding box method [22] to binary images of arbitrary shape. In the gliding box method, an $s \times s$ pixel "gliding box" is initially placed at the upper left corner of the image, and the number of light pixels in the image contained in the gliding box is denoted n_1. The box "glides" over the entire image, moving to all N positions at which it covers at least one pixel of the image, at each location recording the number of light pixels n_i in the image. The sequence of values $\{n_i\}$ for $i \in \{1, 2, \ldots, N\}$ defines a probability distribution $Q_n(M, s)$ which represents the probability that a gliding box of side r contains M light pixels. Lacunarity is defined in terms of the moments $Z_{Q_n}^{(q)}(s) = \sum_{M=1}^{s^2} M^q Q_n(M, s)$ of the distribution $Q_n(M, s)$, by

$$\Lambda(s) = Z_{Q_n}^{(2)}(s)/[Z_{Q_n}^{(1)}(s)]^2.$$

When working with irregularly shaped images such as nuclei, at many positions—in particular, near the periphery of an image—many of the pixels in the gliding box will not be part of the image, but "background" instead. The values of n_i at these positions will be low since there are few pixels of the image within the gliding box at these locations. The net result can be to skew the distribution $Q_n(M, s)$. To correct for this problem, we incorporate a weighting factor w_i, equal to the number of pixels in the gliding box that are part of the image. Define x_i, the weighted number of light pixels in the image contained in the gliding box, by

$$x_i = n_i(s^2/w_i). \tag{2}$$

Weighted lacunarity is defined in terms of weighted moments of the distribution $Q_x(M, s)$ of $\{x_i\}$. The rationale for weighting the moments as well as the number of pixels is that we do not wish a position for which the gliding box contains only a few pixels in the image to contribute as much to the lacunarity as a position for which it contains many pixels. Instead of summing over the number of light pixels, moments of $Q_x(M, s)$ can be calculated by summing over the N positions taken by the gliding box, i.e.,

$$Z_{Q_x}^{(q)}(s) = \sum_{M=1}^{s^2} M^q Q_x(M, s) = \sum_{i=1}^{N} x_i^q/N = \sum_{i=l}^{N} x_i^q \Big/ \sum_{i=1}^{N} 1\,.$$

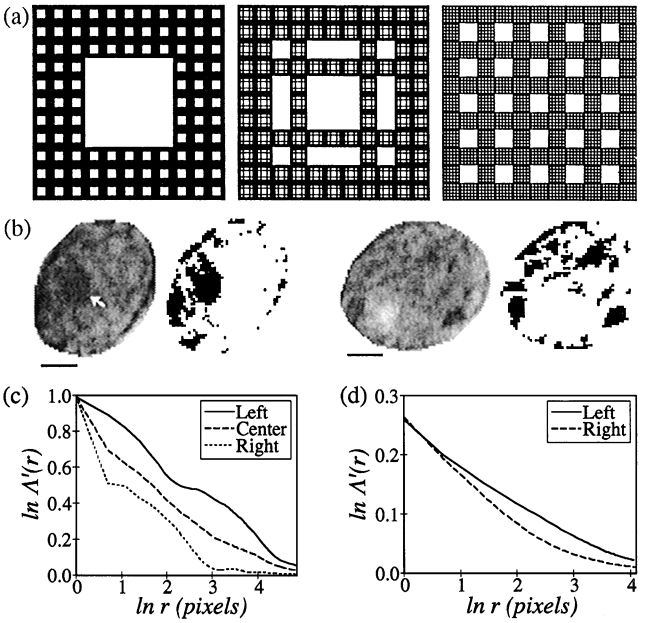

FIG. 2. (a) Two stages of Sierpinksi carpets (inverse images) with Hausdorf dimensions of $\ln 96/\ln 11 \approx 1.90$ and (c) their lacunarity curves. The carpets are denoted in (c) by their positions in (a) as left, right, and center. (b) Two breast epithelial cell nuclei with $D_{\mathrm{MB}} = 2.49$ and $D_S = 2.81$, shown both as gray scale images and thresholded at the first quartile of the intensity histogram, and (d) the lacunarity curves for the two nuclei. The nucleus on the left has a prominent nucleolus, marked by an arrow, and marginated chromatin. The nucleus on the right, despite a nuclear void, has a more uniform distribution of chromatin and is less lacunar. Scale bars are 2 μm.

Weighted moments $Z_{Q_x}^{I(q)}(s)$ are given by

$$Z_{Q_x}^{I(q)}(s) = \sum_{i=1}^{N} w_i x_i^q \Big/ \sum_{i=1}^{N} w_i , \tag{3}$$

and weighted lacunarity is defined as

$$\Lambda'(s) = Z_{Q_x}^{I(2)}(s)/[Z_{Q_x}^{I(1)}(s)]^2. \tag{4}$$

Equivalently, substituting (2) and (3) into (4) yields

$$\Lambda'(s) = \sum_{i=1}^{N}(n_i^2/w_i)\sum_{i=1}^{N} w_i \Big/ \left(\sum_{i=1}^{N} n_i\right)^2 .$$

Weighted lacunarity curves for the Sierpinski carpets and nuclei in Fig. 2 are shown below the images. While these fractals have the same dimensions, they can be distinguished on the basis of the lacunarity curves.

Weighted lacunarity functions were determined for each nucleus thresholded at the third quartile of its intensity histogram, with box side lengths ranging from $s = 2$ pixels (0.269 μm) to $s = 35$ pixels (4.70 μm). Overall means of weighted lacunarity were greater in malignant cases than in benign cases. As is shown in Fig. 3, this difference was significant at virtually all box sizes (maximum p of 0.0505 at $s = 6$), but especially at the two ends of the curve. These ends correspond, respectively, to fine-scale texture, such as chromatin clumping, and to large-scale structures such as nucleoli and voids. Thus, lacunarity complements fractal dimensions in characterizing chromatin texture.

Quantitative nuclear properties, such as dimensionality and lacunarity, whose patient means or standard deviations differ significantly between benign and malignant cases, can be combined in a classificatory scheme to diagnose unknown cases. We used logistic regression to classify cases by jackknife (leave-one-out) analysis, considering a number of models incorporating various combinations of fractal textural features as well as nuclear area. The best performance was achieved by a model predicting diagnosis for a patient on the basis of mean D_S, mean normalized $\Lambda'(35)$, mean area, standard deviation of C, and standard deviation of normalized $\ln \Lambda'(2)$. This yielded a correct diagnosis for all 19 benign cases and 20 of the 22 malignant cases, an accuracy of 95.1%. These findings suggest that measures of self-affinity can be used as clinical parameters to assist cytologists in rendering an accurate diagnosis in FNA of the breast, and potentially in automated instruments for quality assurance in pathology.

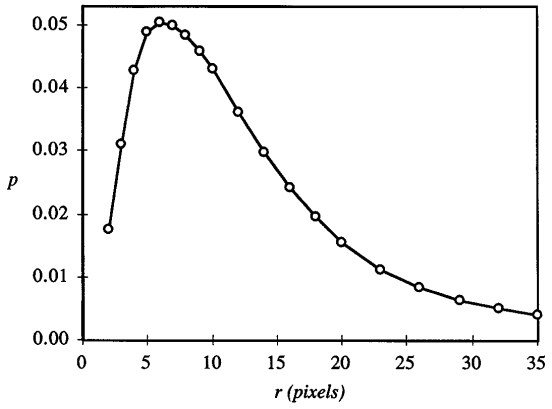

FIG. 3. p value as a function of box size. p is the significance level of the student's t-test comparing mean lacunarities of benign and malignant patients.

We thank C. Benham, S. Wallenstein, C. Bodian, and W. Lou for helpful discussion, M. Sanchez and G. Yang for materials, and D. Burstein and I. Bleiweiss for diagnoses. A. J. E. was supported by NIH MSTP Grant No. GM07280.

*Corresponding author.
Electronic address: einstein@msvax.mssm.edu

[1] L. J. Layfield, B. J. Glasgow, and H. Cramer, Pathol. Annu. **24**, 23 (1989).
[2] Y. Gazit *et al.*, Phys. Rev. Lett. **75**, 2428 (1995); C.-K. Peng *et al., ibid.* **70**, 1343 (1993).
[3] G. A. Losa, G. Baumann, and T. F. Nonnenmacher, Acta Stereol. **11/I**, 335–341 (1992).
[4] G. Landini and J. W. Rippin, Fractals **1**, 326 (1993).
[5] S. S. Cross *et al.*, J. Pathol. **172**, 317 (1994).
[6] H. Banda-Gamboa *et al.*, Anal. Cell. Pathol. **5**, 85 (1993).
[7] T. Jørgensen *et al.*, Cytometry **24**, 277 (1996) .
[8] A. P. Pentland, IEEE Trans. Pattern Anal. Mach. Intell. **6**, 661 (1984).
[9] B. B. Mandelbrot, C. R. Acad. Sci. Ser. A **288**, 81 (1979).
[10] Y. C. Oertel, *Fine Needle Aspiration of the Breast* (Butterworths, Boston, 1987).
[11] G. C. H. Yang and I. I. Alvarez, Acta Cytol. **39**, 55 (1995).
[12] A. J. Einstein *et al.*, J. Microsc. **176**, 158 (1994).
[13] B. B. Mandelbrot, Phys. Scr. **32**, 257 (1985).
[14] B. Dubuc *et al.*, Proc. R. Soc. London A **425**, 113 (1989).
[15] M. Takahashi, J. Theor. Biol. **141**, 117 (1989) proposes a fractal model of metaphase chromosomes, while A. Grosberg *et al.*, Europhys. Lett. **23**, 373 (1993) suggests that long range correlations observed in DNA can be accounted for by a crumpled globule model in which three-dimensional DNA structure is self-similar.
[16] J. P. Rigaut *et al.*, in *Fractals in Biology and Medicine*, edited by G. A. Losa *et al.* (Birkhäuser, Basel, 1997), Vol. 2.
[17] J. P. Rigaut, J. Microsc. **133**, 41 (1984).
[18] D. Hamburger, O. Biham, and D. Avnir, Phys. Rev. E **53**, 3342 (1996).
[19] R. F. Voss, in *Fundamental Algorithms for Computer Graphics*, edited by R. A. Earnshaw (Springer-Verlag, Berlin, 1985).
[20] H.-S. Wu, J. Gil, and A. J. Einstein, Fractals (to be published).
[21] A. L. Goldberger *et al.*, Am. Heart J. **107**, 612 (1984).
[22] C. Allain and M. Cloitre, Phys. Rev. A **44**, 3552 (1991).