Complexity, fractals, disease time, and cancer

W. B. Spillman, Jr.,[1,2] J. L. Robertson,[3,4] W. R. Huckle,[4] B. S. Govindan,[1] and K. E. Meissner[1]

[1]*Virginia Tech Applied Biosciences Center, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*

[2]*Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*

[3]*Center for Comparative Oncology at Virginia Tech, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*

[4]*Biomedical Sciences & Pathobiology Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*

Despite many years of research, a method to precisely and quantitatively determine cancer disease state remains elusive. Current practice for characterizing solid tumors involves the use of varying systems of tumor grading and staging and thus leaves diagnosis and clinical staging dependent on the experience and skill of the physicians involved. Although numerous disease markers have been identified, no combination of them has yet been found that produces a quantifiable and reliable measure of disease state. Newly developed genomic markers and other measures based on the developing sciences of complexity offer promise that this situation may soon be changed for the better. In this paper, we examine the potential of two measures of complexity, fractal dimension and percolation, for use as components of a yet to be determined "disease time" vector that more accurately quantifies disease state. The measures are applied to a set of micrographs of progressive rat hepatoma and analyzed in terms of their correlation with cell differentiation, ratio of tumor weight to rat body weight and tumor growth time. The results provide some support for the idea that measures of complexity could be important elements of any future cancer "disease time" vector.

INTRODUCTION

One of the principal characteristics of biological systems is that they are complex [1]. This complexity allows an evolving optimal functioning within the environment. This fact that has been recognized in systems engineering and is emphasized by the increasing number of engineered nonbiological but biomimetic systems that have appeared over the past few decades [2–4]. Although living systems are adaptive and evolve over time, the mapping of system evolution to chronological time is most often nonlinear [5]. That being the case, additional measures or markers are used to supplement chronological time, when, for example, the state of disease progression is assessed.

If complexity is one of the hallmarks of living systems, then its measure might provide a way to more accurately and quantitatively characterize disease progression. For that purpose, we have applied measures of complexity to a particular type of cancer, rat hepatoma, and looked for signs of correlation between disease state and complexity measures.

Tumors are considered by many to be more complex than the normal tissues from which they are derived. One measure of complexity is morphologic complexity. Tumor cells, by definition, show variations from normal size, shape, and the ratio of nucleus to cytoplasm. Less differentiated, and more malignant tumors, have a higher percentage of cells that are either larger than normal cells or smaller than normal or both. Such cells may have compound or multiple nuclei within a single cell cytoplasm, just as one example of variability and complexity.

The functional evaluation of tumors also shows they exhibit less predictable behavior than normal cells. Tumor cells may utilize both aerobic and anaerobic metabolism to produce energy for cell activities, synthesize unusual or aberrant proteins, or may utilize derepressed portions of fetal/embryonic genome. Many tumor cells show immaturity of the cell surface, a higher degree of cell lability and have a propensity for more mutation, all features not seen in most differentiated adult cells. Some populations of tumor cells resist normal apoptotic signals, and have lengthened lifespan (as a population may be considered immortal), while other tumor cells in the same tumor may die rapidly, being unable to adapt to the conditions of growth characterizing the tumor (hypoxic environment, for example). Although nearly a century of research has sought to identify a "simple" characteristic possessed by all tumor cells, the result of such inquiry has been recognition of the constantly evolving and complex character of all neoplastic cells.

In seeking to apply measures of complexity to cancer in order to determine the correlation of those measures with the disease state, two serious problems exist. First, there are issues in identifying disease state in solid cancer.

"*Tumor grade indicates the level of malignancy and is based on the degree of anaplasia (or deformity in behavior and form) seen in cancerous cells under the microscope… There is still no general clinical agreement on the grading* [6]."

Second, although there are a number of measures of system complexity [7], there is also no agreement on any particular one as having general applicability. The implication of this is that there is no agreement as to what measure of complexity should be considered as an independent variable (what is the *x* axis?) while the tumor grading is poorly defined (what are
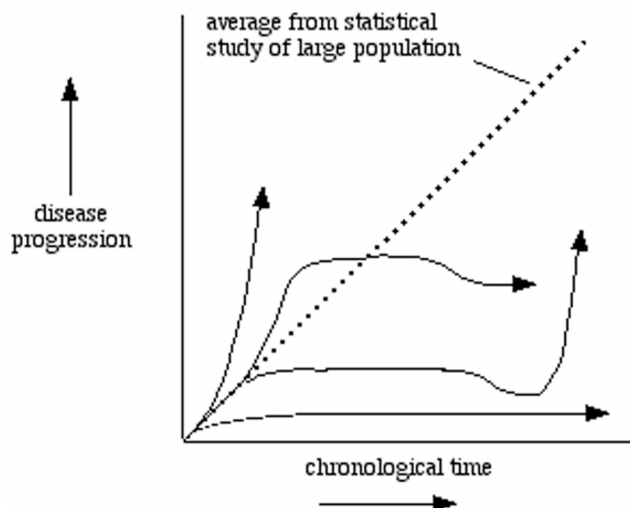
FIG. 1. Disease progression vs chronological time.

FIG. 2. Disease progression vs "disease time."

the values on the y axis?). This makes a meaningful analysis problematic, but the problem can be overcome to a certain extent, as shown below.

Cancer is a complex disease process characterized by the uncontrolled proliferation of abnormal cells. Treatment plans for most cancer patients are based on the interpretation of diagnostic biopsies and clinical staging of disease (primarily by evaluation of primary and distant tumor sites, including lymph nodes). For practical reasons, treatment plans are formulated on "average" outcomes, derived from the results of clinical studies of similar cancers, and not geared to individual patients. As such, most patients can be expected (based on epidemiological data) to respond in a predictable manner to therapy. Many will respond, but some will not, since their individual pattern of disease progression does not fit "the average" as shown in Fig. 1. There have been striking instances in which intensively managed *individual* patients attain significantly better disease control than those managed according to methods based on average outcome (the winner of the 2003 Tour de France bicycle race, Lance Armstrong, with metastatic testicular cancer, for example). Of course, some patients do not do as well as "the average," have poorer response to therapy, and succumb sooner.

Similar neoplasms grow and progress differently in different individuals. Underlying reasons for this differential pattern of growth include the type(s) of mutations leading to initial cancer cell growth, success of cancer cells in developing vasculature and support stroma, patient immune and inflammatory response to cancer cells, patient and tumor nutrition, and the ability of the cancer cells to evolve resistant forms in response to attack. The rate of growth of neoplasms follows a relatively predictable biological, but not chronological, pattern in each individual. We identify this predictable biological growth pattern as the disease state reflecting different stages in the growth, differentiation, and progression of the cancer and measured by a hypothetical "disease time" as shown in Fig. 2.

Therapy is designed to kill cancer cells while leaving normal cells intact. In general, the type, duration and amount of therapy are based on the histologic assessment of tumor type.
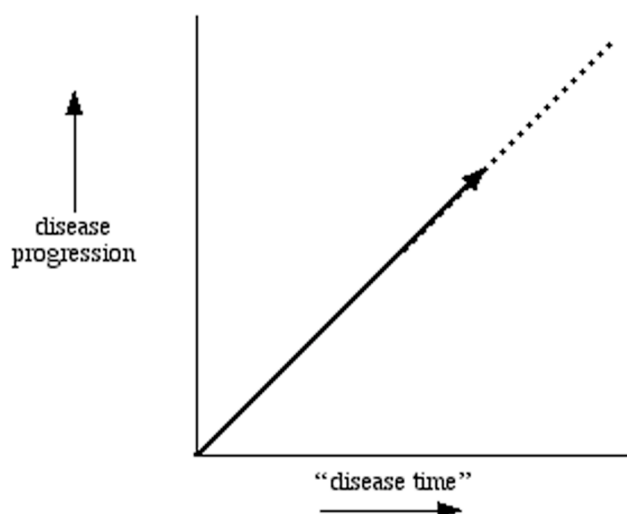
From this assessment, a prediction of future growth is made based on past experience with this tumor morphology. Therapy is not usually based on (or monitored by) repetitive reevaluation of tumor morphology, ongoing assessment of tumor gene expression, or other measures of cancer "disease time." Increasing the predictive value of the surgical biopsy and additional incremental evaluation of cancers may provide better control of cancer growth and better patient outcomes through individualizing therapeutic planning and monitoring.

Pathologists evaluate a variety of microscopic features when analyzing tumor biopsies. Features such as cell size, nuclear size, nuclear shape, ratio of nuclear to cytoplasmic size, subcellular organelle occurrence and placement (nucleoli, for example) are commonly evaluated. The organization of neoplastic cells into tissue architecture such as ducts and glands is considered important. The presence or absence of mitotic figures is typically taken as a measure of cell growth fraction and predictive of ongoing cell division. The general shape of the neoplastic cell population and degree of invasiveness may be analyzed as additional criteria indicating benign or malignant behavior.

Pathologists commonly evaluate the morphology in many different areas of the surgical biopsy to arrive at a definitive morphologic diagnosis that will be used to define the potential future growth of the neoplasm and therapy for the disease. In most cases, the least differentiated and/or most poorly organized area of the tumor is taken as being the area most predictive of future growth.

The experience of the pathologist is the key to arriving at an accurate diagnosis. This experience is gained from years of examining surgical biopsies, studying and comparing morphologic features, and relating these images to eventual case outcomes. However, even highly experienced pathologists examining the same biopsies may disagree on the diagnosis, suggesting fundamental differences in perception, visualization, and utilization of their recalled images [8].

In recent times, as diagnosis and treatment of cancer have advanced, there has been much emphasis on the use of quan-

titative and semiquantitative methods in analyzing cancer micrographs in order to quantify the exhibited morphological features. However, the tremendous diversity seen both in the type as well as in the other characteristics of cancer makes this task difficult. In particular, the disease progression rate shows wide variation even between experiments done under similar circumstances [9]. As a result, the usual statistical averaging using chronological time as a parameter cannot be expected to lead to reliable predictions. It is thus highly desirable to identify a suitable variable, or set of independent variables, which would help to characterize cancer disease state accurately in terms of a suitable "disease time" vector.

Many cancer markers have been identified over the years [10], but no combination of them has yet allowed unambiguous quantitative determination of disease state. Recent advances in bioinformatics have uncovered a number of genomic markers that are related to specific cancers. These markers, whose significance is only beginning to be fully understood, offer great potential for use in the creation of a "disease time" vector.

There are other potential markers that can be derived from the evolving sciences of complexity. The study of complex adaptive systems (which includes all living systems) took root in the second half of the 20th century and has been progressing ever since. As in many new fields, some definitions have yet to be agreed upon, and in this case, "complexity" itself is one of those. Properties associated with complexity include diversity, Shannon entropy, the shortest algorithm needed to describe the system, the algorithm that takes the shortest time to describe the system, fractal dimension, percolation, emergent behavior and others [7]. Complexity is one of those things that everyone can identify but which is difficult to define. An example is found in the transition between water and ice. Both water and ice appear to be simple, rather than complex systems. At the transition between the two, there will be a diversity of ice masses floating in water. A maximum in complexity will occur somewhere in the region between the more simple uniform phases. If one considers the irreducible amount of information required to describe the water-ice system, the maximum amount of information is required for the water phase (mass of a single water molecule, as well as position, momentum and spatial orientation of each water molecule must be specified), while the minimum amount of information is required for the ice phase (total mass, center of mass, total momentum, and lattice parameters must be specified). In terms of the information required to describe a system, we hypothesize that:

A system is maximally complex when the rate of change of the irreducible amount of information required to describe that system with respect to some parameter or parameter set is an extremum.

This means that the absolute value of the slope of the amount of irreducible information required to describe the system versus some parameter reaches a maximum (when the amount of information is changing from a smaller to a larger value, e.g., from ice to water, or the information required to describe it is changing from a larger to a smaller value, e.g., water to ice) when the system is maximally complex. We will consider this hypothesis in the light of our results later, but it also has some relevance to one of the measures of complexity we consider, the fractal dimension of a micrograph.

The fractal dimension [11] is related to the space filling nature of an object, with a two dimensional object such as a square having fractal dimension $D = 2$ and a one dimensional object such as a straight line having fractal dimension $D = 1$. Squares and straight lines are not generally considered to be complex, so one would expect a maximum in complexity to occur for objects having fractal dimensions somewhere between integral values in analogy to finding maximum complexity at the critical point at the phase transition between water and ice, i.e., near $D = 1.5$ for a two-dimensional embedding space.

The other complexity measures we use are both related to percolation. Percolation [1,12] is a measure of the connectivity of a region of space. If at least one unbroken path exists from the bottom to the top of a region in 2 dimensions, the region is said to percolate. Percolation is often described in terms of a "forest fire" model [1]. In this model, a finite lattice is partially populated by "trees" or occupied sites. At time step 1, all the trees on the bottom row of the lattice are set on "fire." At the next time step, the trees that were burning are replaced by "ash" and adjacent unburned trees are set on fire. This process continues until the fire burns out. At that time, lattice sites can either be unoccupied, contain ash or contain an unburned tree. The parameters associated with the process are the percolation time, t_p, the fractional percolation, F, and the total percolation time, T. In this model, the percolation time, t_p, or the number of time steps required for the fire to reach the top of the lattice divided by the number of rows in the lattice, is infinite if no complete path exists from the bottom to the top of the lattice. The fractional percolation, F, is the number of unburned trees divided by the sum of unburned and burned trees. Finally, the total percolation time, T, is the number of time steps required until the fire goes out divided by the number of rows in the lattice. Percolation for an image can be characterized by these parameters if the image is considered to be a two dimensional lattice in which each pixel location represents a lattice site and a site is said to be occupied if its associated pixel is some specified color. A more detailed description of how the fractal dimension and percolation measures are calculated are provided below.

In this paper, we first describe the particular cancer analyzed, rat hepatoma, how the samples were collected, how the slides were prepared and how the images were taken. Next, the standard parameters used to characterize that data set are discussed. The complexity measures of image morphology are then covered in some detail followed by the analysis procedure, i.e., preprocessing of images, conversion of images to data files, calculation of fractal dimension, fractional percolation, total percolation time and an estimate of the irreducible amount of information needed to describe the images. The results of this analysis are then presented including a linear correlation matrix relating all pertinent variables. Finally, the results are discussed, their implications analyzed and the future direction of this research considered.

DATA SET

Female Buffalo rats (8 weeks old) were obtained from a commercial vendor (Harlan Laboratory Animals, Dublin,
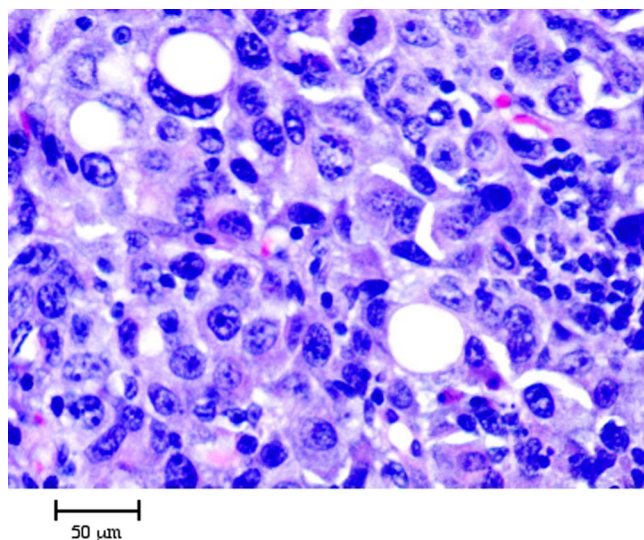
FIG. 3. (Color online) Image 1.

FIG. 4. (Color online) Image 2.

VA), housed singly in polycarbonate cages, fed a commercial, nutritionally complete rodent diet (Harlan Rat Chow), and allowed access to public water *ad libitum*. Prior to the initiation of this work, a protocol for experimentation was submitted and approved by the Institutional Animal Care and Use Committee. During the conduct of experimentation, animals were treated in accordance with guidelines for care and use of laboratory animals as promulgated by AALAC. Work was monitored by the University Veterinarian. Animal use was consistent with highest acceptable standards of humane care and use in experimentation.

Morris 7777 transplantable hepatomas were raised in tissue culture and then an aliquot of 10^7 cells was injected aseptically in the subcutis anterior to the left hind leg of selected rats. Tumor growth was subsequently monitored on a daily basis by gentle palpation. Groups of rats were sacrificed at weekly intervals for 8 weeks, tumors were gently exposed, photographed *in situ*, and then collected for further study. Pieces of tumors were fixed in 10% buffered formalin solution for a minimum of 48 hours, dehydrated after fixation in a series of graded alcohols and xylene, and then infiltrated with paraffin polymer. Sections were cut at 3 microns, rehydrated, and then batch stained on an automated stainer (Shandon Southern Corp) with hematoxylin-eosin stains. Digital photographs of stained four micrometer tumor sections were acquired using a Nikon Eclipse E600 photomicroscope, DXM 1200 Digital Camera, and processed with Nikon ACT-1, version 2 proprietary image capture software. All sections were photographed at original magnification $40\times$. Images were stored as medium compression jpeg files (800×600 pixels). The data set analyzed consisted of 9 images displaying different degrees of cell differentiation. These images are shown in Figs. 3–11 as described below.

Figures 3 and 4: disorganized early growth. Tumor cells show marked variation in cell and nuclear size and shape. Note the presence of enlarged nuclei, taken to be indicative of abortive mitotic events, and nuclear hyperchromasia. Some degenerate tumor cells (arrows) are present and there is minimal inflammatory cell infiltrate.
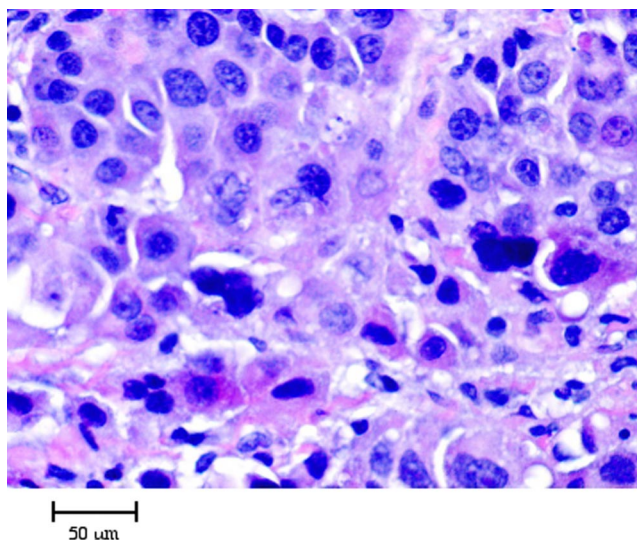
Figures 5–8: tumor maturation is evident with less cellular pleomorphism (size and shape variation), decreased number of mitotic indicators, and early formation of cell cords typical of differentiated hepatomas.

Figures 9 and 10: cell cords are evident in these micrographs, indicating further tumor organization. Several small areas of necrosis (arrows) are present in each figure, and individual necrotic cells are scattered throughout.

Figure 11: well differentiated mature tumor, showing pseudoacinar organization and cells arrayed in cords.

STANDARD PARAMETERS

For each of the images, there was associated data that is related to, but does not completely define, the state of disease. This data was the tumor growth time from incubation in days, the tumor weight and the rat weight in gram weights at the time the tumor was harvested. Table I shows the incu-
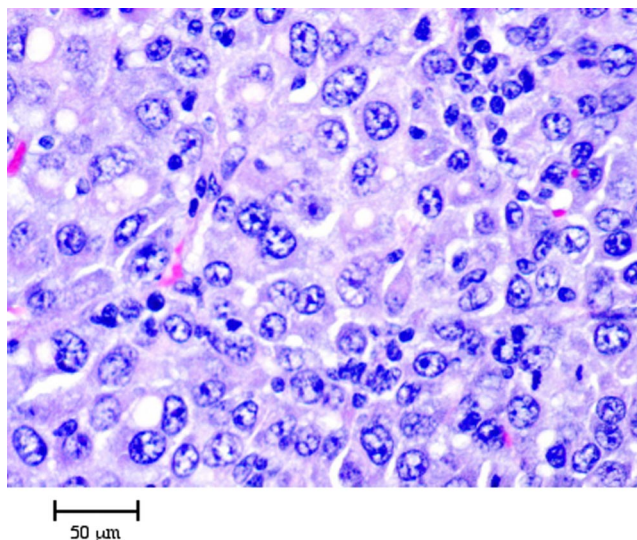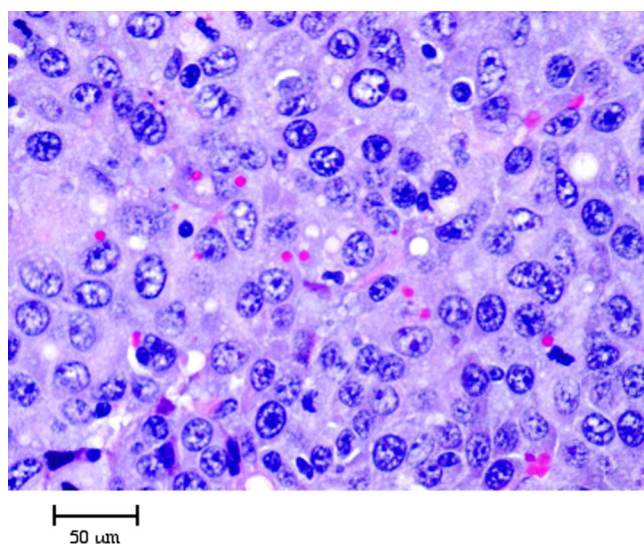

FIG. 5. (Color online) Image 3.

FIG. 6. (Color online) Image 4.

FIG. 8. (Color online) Image 6.

bation time, tumor weight, rat weight and tumor weight/rat weight ratios that correspond to each of the images.

COMPLEXITY MEASURES OF IMAGE MORPHOLOGY

There is a long history of using image analysis to determine the morphology of a system associated with an image [13]. Recently, techniques based on concepts from the field of artificial life [13] have been used for image analysis [14]. The measures of complexity that we use involve fractal dimension and percolation. We will consider each of these in turn.

One factor making determination of the extent of cancer spread difficult is that the typical malignancy has a very irregular, nonsmooth boundary. A large number of experimental studies done on different types of cancer have shown that these irregularities are present over a range of length scales
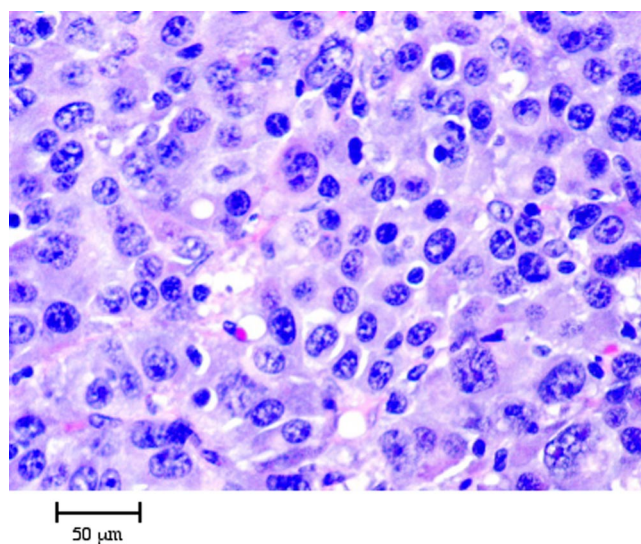
(up to cellular dimensions), implying that the tumor boundary has a *fractal* nature, and could be characterized by a *fractal dimension* [15–17] This means that the tumor invades the available space in a nonuniform way, and has an effective dimension less than the embedding topological dimension (which is 3 for most tumors). Thus, the fractal dimension is a very useful quantity in characterizing the state of growth of the tumor. In general, the roughness of the interface between the tumor and nontumor region is an indicator of whether the tumor is likely to become infiltrative or not. Tumors whose interfaces are very rough are seen to be more aggressive, while ones with relatively smooth boundaries are less likely to be highly infiltrative. The fractal dimension is closely related to the roughness of the interface, which makes it closely correlated to the growth characteristics of the tumor. There has been a considerable amount of research carried out on relating fractal dimension to the presence or absence of cancerous tissue and a number of extensive recent review
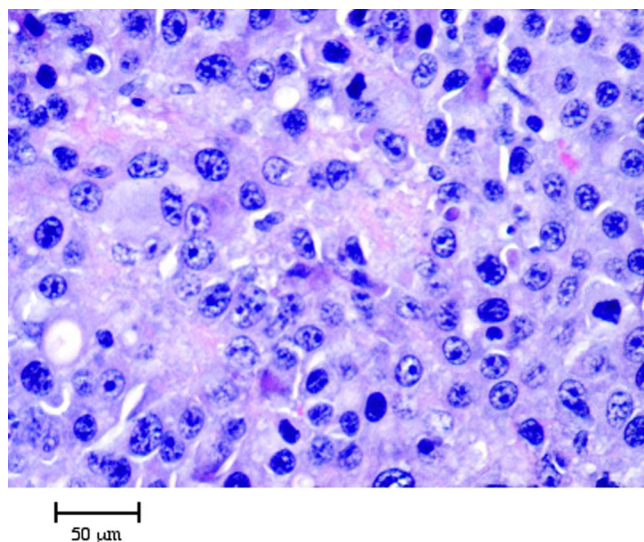

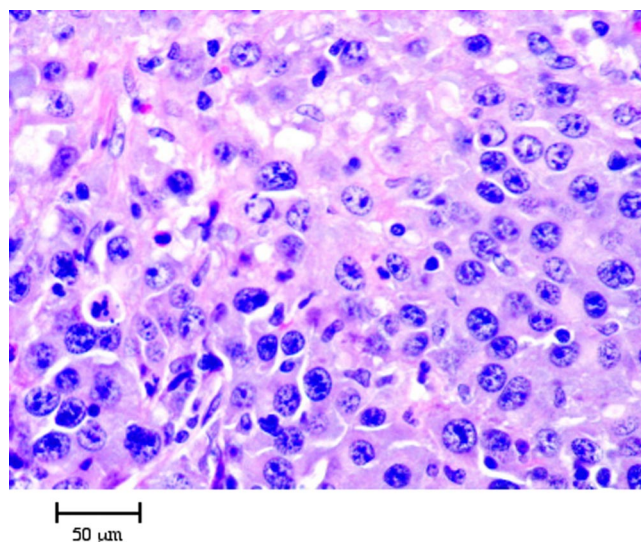
FIG. 7. (Color online) Image 5.
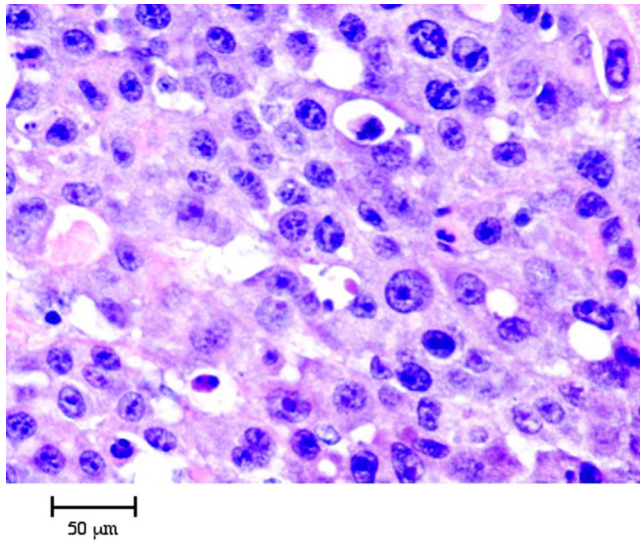

FIG. 9. (Color online) Image 7.
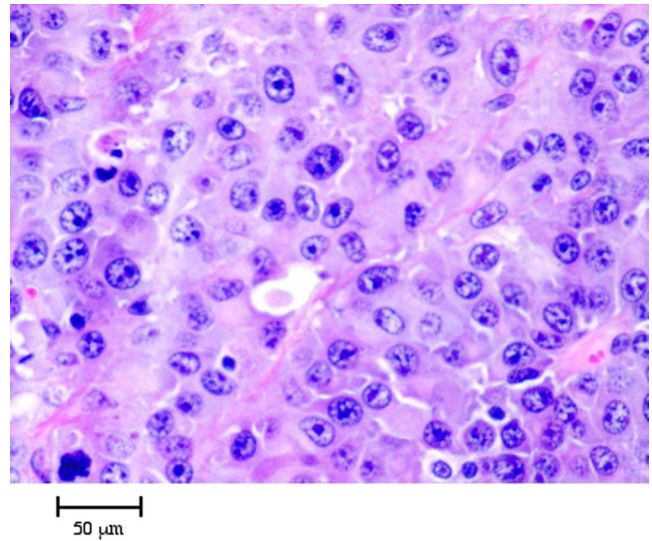
FIG. 10. (Color online) Image 8.

FIG. 11. (Color online) Image 9.

articles are available [15–17]. Changes in tumor vascular architecture have been monitored via determination of the fractal dimension of the associated images [18]. There is a growing but not universal belief that the fractal dimension of cancer images is an important marker that relates to the disease state:

> "*There are strong theoretical reasons for using fractal geometry in measurements of biological systems…*" [16].
> "*If carefully applied, fractal methods may someday have a significant impact on our understanding of challenges in treatment delivery and diagnosis of cancer.*" [19].
> "*…fractal analysis applied on digitized cell shapes is a reliable method for cell complexity measurement that can be used alone or as an additional parameter along with morphometrical measurements both in routine work and research.*" [20].

The concept of a fractal dimension to describe structures which look the same at all length scales was first proposed by Mandelbrot [11]. Although in strict terms, this is a purely mathematical concept, there are many examples in nature that closely approximate a fractal object, though only over particular ranges of scale. Such objects are usually referred

TABLE I. Tumor growth times and tumor weight/rat weight ratios corresponding to each image.

Image No.	Growth time (days)	Tumor weight/rat weight
1	7	0.000
2	7	0.000
3	7	0.000
4	14	0.007
5	14	0.002
6	14	0.012
7	21	0.026
8	35	0.021
9	35	0.044

to as *self-similar* to indicate their scale-invariant structure. In simple terms, the common characteristic of such fractal objects is that their length (if the object is a curve, otherwise it could be the area or volume) depends on the length scale used to measure it, and the fractal dimension tells us the precise nature of this dependence. A rigorous mathematical definition of the fractal dimension, D (also called Haussdorff dimension), of such an object is usually expressed through the relation

$$D = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)}, \qquad (1)$$

where D is the fractal dimension, ε is the length of one side of a hypercube which has the same dimension, n, as the space in which an object is embedded and $N(\varepsilon)$ is the minimum number of hypercubes of dimension ε required to completely cover the object. Since one cannot achieve the limit of ε going to 0 in numerical calculations, $\ln N(\varepsilon)$ and $\ln(1/\varepsilon)$ are calculated over a range of scales using hypercubes varying in size from the smallest single cube required to cover the object to cubes at the resolution limit of the set of points chosen to represent the object [when this limit is reached $N(\varepsilon)$ equals the number of points and remains constant for all smaller values of ε]. These values are then plotted with the slope representing an "instantaneous fractal dimension." In practice, one would like to find a range of scale of an order of magnitude or more over which the slope is constant. The object will then be said to have a fractal dimension D = slope and self-similarity over that range. Depending upon the range over which the slope is calculated, D can vary widely. This technique is called the box counting algorithm (BCA).

Most calculations of fractal dimension of digital images simply convert the image to black and white according to some protocol, apply the BCA to the set of pixel positions represented by the black pixel data set, find a range where the slope is constant, and conclude that the image has a fractal dimension equal to the slope within the range where the

slope is constant. A closer examination of the BCA, however, provides more information about the morphology of the image than this procedure suggests. If an object is represented by an array of points embedded in an n dimensional space, then we define

$$N(\varepsilon) = N_0(\varepsilon) - N_\varepsilon, \qquad (2)$$

where $N(\varepsilon)$ is the number of hypercubes of size ε required to cover the set of points, $N_0(\varepsilon)$ is the number of hypercubes of size ε required to completely fill the smallest single hypercube that can completely contain all the points, and N_ε is the number of hypercubes of size ε (in the smallest single hypercube that can completely contain all the points) that are empty.

If V_0 is the volume of the smallest single hypercube that can completely contain the set of points and V_ε is the volume of all the voids within the object of dimension ε or larger, then in an approximate sense

$$N(\varepsilon) \approx V_0/\varepsilon^n - V_\varepsilon/\varepsilon^n = V_0(1/\varepsilon)^n(1 - V_\varepsilon/V_0) \qquad (3)$$

and

$$\ln N(\varepsilon) \approx \ln(V_0) + n \ln(1/\varepsilon) + \ln(1 - V_\varepsilon/V_0). \qquad (4)$$

An examination of Eq. (4) indicates that if there are no voids in an object with dimension smaller than some given ε, then the first and third terms on the right-hand side of Eq. (4) are constant for values of ε smaller than this and the slope of the curve generated by the BCA for scales ε and smaller should be close to the dimension of the space in which the set of points is embedded. Deviations of the slope from the dimension of the embedding space at a given scale are related to the distribution of voids at that scale. This leads to the interesting speculation that:

The curve produced by the BCA for a set of points is related to the distribution of voids within that set of points with the corollary that the curve produced by the BCA on the complimentary set of points should be related the distribution of points themselves.

Our simple analysis of the BCA has two implications for the analysis of cancer micrographs: (1) the scale at which morphological differences between images occur might be evident from the scale at which the BCA curves for images deviate from one another, and (2) the BCA curve for the complimentary set of points (the locations of the white pixels) and its fractal dimension might be more correlated with the tissue morphology than the fractal dimension of set of points itself. Due to the latter consideration, the fractal dimension of both the black pixel set and the white pixel set was processed for each image.

In implementing the BCA in a computer algorithm, we followed a recently reported innovative technique that minimizes memory requirements [21]. In addition, each set of data points was scaled so that each point was replaced by (0.98 times its value)/(the maximum x-axis value in the set) +0.01. The algorithm was validated on data sets representing a linear object (sine curve), a two-dimensional object (filled in square) and a complex object (Henon Map).

For the percolation measurements, a computer program was developed that implemented the commonly utilized forest fire model [1]. In this program, pixels were considered to be adjacent on the square lattice if they were nearest neighbors or nearest diagonal neighbors (8 possible total adjacent locations). In previous work devoted to stochastic modeling of molecular self-assembly [22], it was found that fractional percolation and total percolation time were sensitive indicators of structural phase transitions in the model films, and might provide an indication of the same sort of transitions, if they exist, as in disease progression.

The computer algorithms were written in the FutureBasic programming language and run on a Macintosh G5 computer using the OS 10.2 operating sysem.

ANALYSIS PROCEDURE AND RESULTS

The rat hematoma images were preprocessed using Adobe Photoshop 7.0 image processing software. The processing proceeded in the following way.

(1) In order to reduce computation times, the color jpeg images were rescaled from an original size of 800 $\times$ 600 pixels to a reduced size of 468 $\times$ 351 pixels using bicubic pixel interpolation. This reduction resulted in image degradation at the highest resolution and limited the range over which meaningful calculations could take place. The physical distance captured by each image from left to right
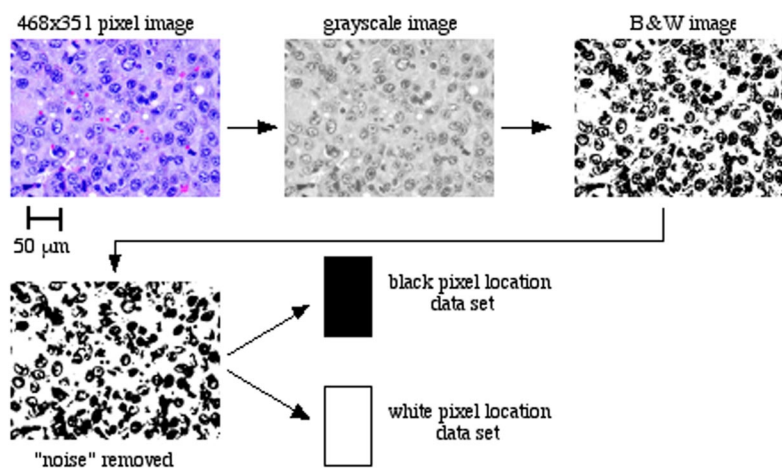

FIG. 12. (Color online) Image preprocessing.

TABLE II. Values of calculated parameters for the 9 rat hepatoma images.

Image No.	D_b	D_w	F_b	F_w	T_b	T_w	I_{min} (kb)
1	1.664	1.776	0.105	0.984	0.211	1.114	269.2
2	1.631	1.804	0.091	0.995	0.162	1.051	241.8
3	1.611	1.822	0.166	0.970	0.601	1.154	272.0
4	1.626	1.811	0.094	0.964	0.205	1.125	280.1
5	1.600	1.823	0.190	0.987	0.368	1.088	260.3
6	1.625	1.804	0.069	0.986	0.182	1.040	260.5
7	1.617	1.812	0.023	0.983	0.083	1.142	256.1
8	1.635	1.793	0.216	0.988	0.550	1.083	246.1
9	1.634	1.799	0.091	0.956	0.285	1.259	226.6

FIG. 13. Fractal dimension of the black pixel sets vs image order for the 4 pathologists.

was 0.400 mm, so that the distance between the centers of individual pixels was 0.854 μm.

(2) The color jpeg images were converted to gray scale.

(3) The gray scale images were converted to black and white using a threshold level of 0.5$\times$ maximum intensity.

(4) The black and white images were filtered to remove stray pixel "noise" using the "dust removal option" with a 2 pixel radius.

(5) The black and white images were then converted to two data sets containing the locations of the black pixels and white pixels, respectively.

(6) The original color jpeg images were compressed using the standard gzip algorithm and the files sizes, I_{min}, were recorded as an indication of the irreducible amount of information required to describe those images.
This process is shown in Fig. 12.

Once the data sets had been obtained, the BCA was used to generate $\ln N(\varepsilon)$ versus $\ln(1/\varepsilon)$ curves for both the black pixel and white pixel data sets for each of the 9 images. The starting value was $\varepsilon = 1$ and the ending value was $\varepsilon = 0.0015$ so that ε ranged over almost 3 orders of magnitude. The step reduction factor for the ε's was 0.81. It was found that the curves were linear ($r^2 \geq 0.999$) in the region $0.003 \leq \varepsilon \leq 0.052$ (14 points, $\sim$ one order of magnitude) for all 18
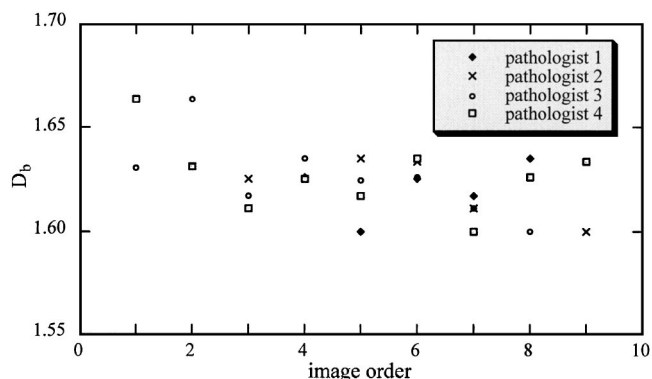
curves. The values of the slopes of the curves in that region were identified as the fractal dimensions and complimentary fractal dimensions of their associated images.

The black pixel and white pixel data sets were used as input to the percolation analysis computer program and the fractional percolation, F, and the total percolation time, T, were calculated. The calculated values for each of the 9 images are given in Table II.

The fractal dimension, D, percolation parameters F and T and the irreducible amount of information, I_{min}, needed to describe the images were considered to be independent variables in this analysis. The other independent variable is of course the tumor growth time, t. In order to determine the correlation of the independent variables to disease state, they needed to be compared with dependent variables that are commonly used in the assessment of disease state, cell differentiation, d, and ratio of tumor weight to rat weight, R. Two of the variables have already been presented in Table I, i.e., the tumor growth time and the tumor weight to rat weight ratio. The second dependent variable is the degree of cell differentiation in the images. In order to get a quantifiable measure of this variable, 4 pathologists were asked to independently arrange the images from the lowest degree (1) of cell differentiation (undifferentiated) to the highest degree

TABLE III. Cell differentiation (1=undifferentiated; 9=highly differentiated) image orders. Pathologist 1: Dr. John L. Robertson, Center for Comparative Oncology. Pathologist 2: Dr. Robert Maronpot, National Institute of Environmental Health Science. Pathologist 3: Dr. Gerry Long, Eli Lilly Inc. Pathologist 4: Dr. Kurt Zimmerman, Virginia-Maryland Regional College of Veterinary Medicine.

Image No.	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Pathologist average
1	1	1	2	1	1.25
2	2	2	1	2	1.75
3	3	7	7	3	5.00
4	4	8	6	8	6.50
5	5	9	8	7	7.25
6	6	3	5	4	4.50
7	7	4	3	5	4.75
8	8	5	4	6	5.75
9	9	6	9	9	8.25

TABLE IV. Linear correlation matrix of relevant parameters. D_b=fractal dimension of the black pixel sets. D_w=fractal dimension of the white pixel sets. F_b=fractional percolation of the black pixel sets. F_w=fractional percolation of the white pixel sets. T_b=total percolation time of the black pixel sets. T_w=total percolation time of the white pixel sets. I_{min}=irreducible amount of information to describe the image. d=cell differentiation from the average image rankings of the 4 pathologists. R=tumor weight/rat weight. t=tumor growth time.

	D_b	D_w	F_b	F_w	T_b	T_w	I_{min}	d	R	t
D_b	1	0.97	0.18	0.03	0.24	0.05	0.11	0.54	0.06	0.04
D_w	0.97	1	0.10	0.14	0.20	0.02	0.21	0.48	0.14	0.18
F_b	0.18	0.10	1	0.12	0.88	0.14	0.04	0.26	0.27	0.15
F_w	0.03	0.14	0.12	1	0.14	0.85	0.01	0.62	0.48	0.31
T_b	0.24	0.20	0.88	0.14	1	0.12	0.05	0.31	0.11	0.19
T_w	0.05	0.02	0.14	0.85	0.12	1	0.30	0.52	0.65	0.45
I_{min}	0.11	0.21	0.04	0.01	0.05	0.30	1	0.20	0.68	0.65
d	0.54	0.48	0.26	0.62	0.31	0.52	0.20	1	0.57	0.63
R	0.06	0.14	0.27	0.48	0.11	0.65	0.68	0.57	1	0.88
t	0.04	0.18	0.15	0.31	0.19	0.45	0.65	0.63	0.88	1

(9) of cell differentiation (well differentiated). The image ordering by the individual pathologists and the average order of the images is given in Table III. The fractal dimensions of the images versus image order for all 4 pathologists is shown in Fig. 13, where no obvious correlation can be observed. However, the measure of degree of the cell differentiation of each of the images was chosen to be its average image order position and this was used in the subsequent analysis.

The final step in the analysis was to calculate the linear correlation matrix (r values) for all of the variables determined, i.e., fractal dimension of black pixel data sets, D_b, fractal dimension of white pixel data sets, D_w, fractional percolation of black pixel data sets, F_b, fractional percolation of white pixel data sets, F_w, total percolation time of black pixel data sets, T_b, total percolation time of white pixel data sets, and the irreducible amount of information, I_{min}, needed to describe the images, degree of cell differentiation, d, the ratio of tumor weight to rat weight, R, and the tumor growth
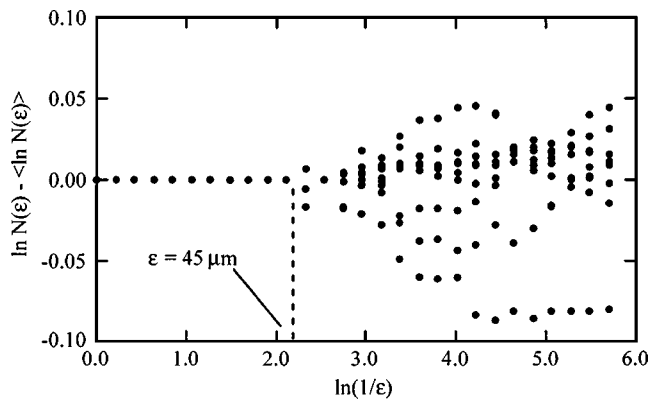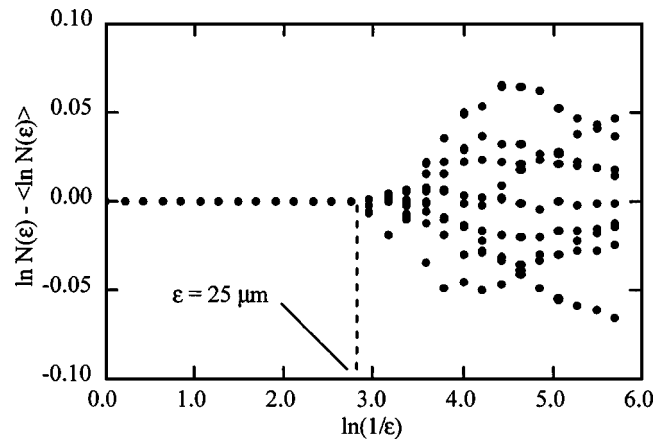

FIG. 15. Deviation of all white pixel BCA curves from their average.

time, t. An r value for 2 parameters of 0.585 for 9 points means there is only a 10% chance that two variables are unrelated while an r value of 0.670 means there is only a 5% chance they are unrelated [23]. The linear correlation matrix is given in Table IV. Figures 14 and 15 show the deviation of the BCA generated curves for the black and white pixel sets from their averages. In Figs. 16 and 17, cell differentiation versus fractal dimension of the black pixel data sets and white pixel data sets are presented with linear and quadratic fits, respectively. Figure 18 shows the first derivative of the irreducible information required to describe the original images (as estimated by their compressed file size using the gzip differentiation.

SUMMARY, DISCUSSION, AND CONCLUSIONS

In carrying out this research, our primary goal was to investigate whether measures of complexity had the potential to be used as components of a hypothetical cancer "disease time" vector (multiple marker) that would provide an accurate quantifiable measure of disease state. This is a difficult problem, since disease state (tumor grading for solid tumors) consists of broad categories, and physician skill and experi-

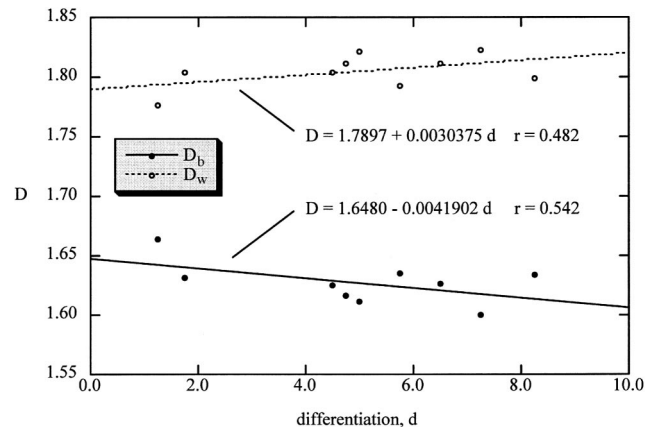FIG. 14. Deviation of all black pixel BCA curves from their average.

FIG. 16. Fractal dimension vs cell differentiation for black and white pixel sets with Linear fits.
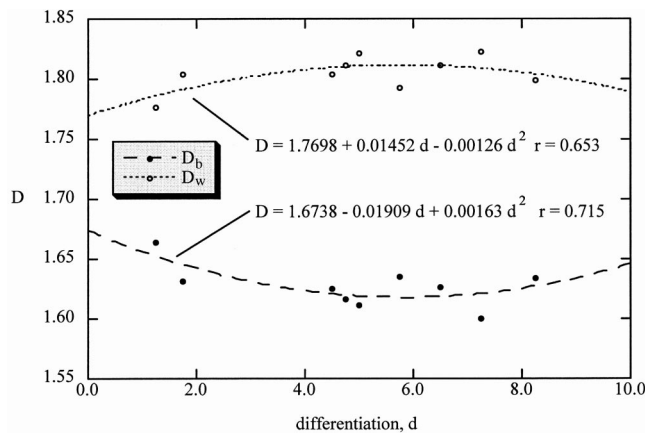
FIG. 17. Fractal dimension vs cell differentiation for black and white pixel sets with quadratic fits.

ence play a large role in determining the accuracy of the grade to which any given tumor is assigned. If one is attempting to do correlations, this results in an ill-defined "*y* axis" is ill-defined. In our case, we examined 2 dependent variables that are usually associated with disease state in solid tumors, tumor cell differentiation, *d*, and the ratio of tumor weight to animal weight, *R*. Of the two, the most uncertain is cell differentiation, which is a somewhat qualitative measure of tumor morphology utilizing broad categories. When asked to arrange the 9 tumor micrographs in order of cell differentiation, 4 experienced pathologists came up with 4 different orderings. This illustrates the difficulty in quantifying cell differentiation as indicated in Table III and shown in Fig. 13. Our measure of the cell differentiation for each image was chosen as the average place in the order considering all 4 pathologists. Although approximate, this allowed us to determine the amount of correlation between the dependent variables, *d* and *R*, and the independent variables: tumor growth time, *t*, and measures of complexity fractal dimension, *D*, fractional percolation, *F*, and total percolation time, *T*.

A number of hypotheses were tested in this research. First, in a consideration of the box counting algorithm (BCA), analysis suggested that differences in morphology between images might be manifested as differences between the curves generated by the BCA, allowing the scale at which significant features might differ between images to be determined. A mathematical analysis of the BCA also suggested that its application to the white pixel location data set would provide information about the black pixel location data set and vice versa, with the implication that the fractal dimension measures of the white pixel data set would correlate better with the dependent disease state variables than the fractal dimension measures of the black pixel location data set. For that reason, for each image, *D*, *F*, and *T* were determined for both data sets. A final supposition was that maximum complexity in terms of fractal dimension would occur somewhere in the middle region between $D=1$ and $D=2$, i.e., a line and a filled in two-dimensional object such as a rectangle. We will examine each of these suppositions in turn while noting that the linear correlation coefficient between tumor growth time and tumor weight/rat weight was $r = 0.88$, and the linear correlation coefficient between tumor growth time and cell differentiation was $r=0.64$. These results indicate that (1) correlations between the independent complexity based variables and the dependent variables should be greater than 0.6 to be of significance, and (2) we can have some measure of confidence in our approximate measure of cell differentiation.

The differences between the BCA curves and their averages are shown in Fig. 14 for the black pixel location data sets and in Fig. 15 for the white pixel location data sets. As can be seen, there are reasonably clearly defined scales at which the curves suddenly diverge. This indicates that at smaller scales, the number of boxes required to completely contain the individual pixel sets becomes different for the 9 images and implies that analysis of the images at larger scales would probably not produce useful results. In the case of the black pixel location BCA curves, this deviation occurs at ~ 45 μm, while for the white pixel locations, the deviation occurs at ~ 25 μm. If the BCA curves for the white pixel location data sets contain information about the cell morphology and the BCA curves for the black pixel location data sets contain information about the morphology of the spaces between the cells, this implies that deviations in the images from each other in terms of the dimensions of the spaces between the cells begin to occur at a larger scale than deviations in the images from each other in terms of the dimensions of the cells themselves. In effect, the analytical process is able to quantify complexity (fractal) measures of benign
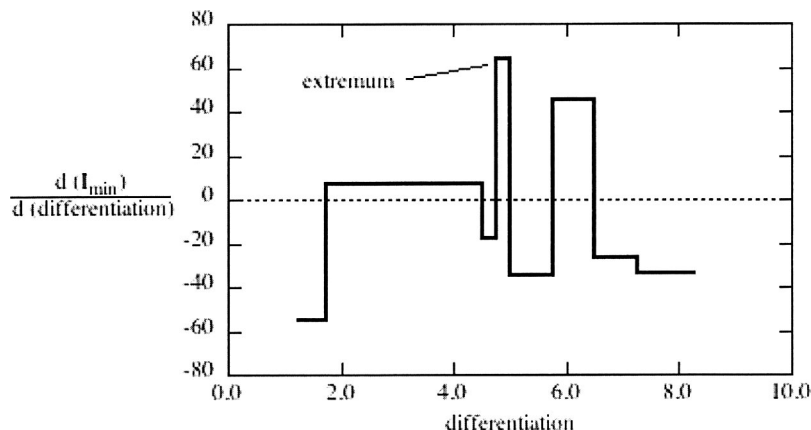

FIG. 18. Rate of change of I_{min} vs cell differentiation.

and malignant morphology. The data indicate significant differences in white pixel (cell morphology) and black pixel (intercellular space organization) data. This is not unexpected, since tumor cell "packing" varies widely if tumors are attempting to replicate tissue architecture (that is, to become more orderly and therefore less complex) or fail to do so (are more malignant and undifferentiated). The real value of these measures is that they provide a quantifiable attribute of cell organization or disorganization, something that is difficult for pathologists to do.

Differences between the black pixel location data sets and the white pixel location data sets were found when examining the linear correlation data presented in Table IV but these were not consistent. Significant correlations between the dependent variables and independent complexity variables ($r > 0.6$) occurred in four cases: F_w versus d ($r=0.62$), T_w versus R ($r=0.65$), I_{min} versus R ($r=0.68$), and I_{min} versus t ($r=0.65$). Four other correlations approached the 0.6 value: D_b versus d ($r=0.54$), T_w versus d ($r=0.52$), D_w versus d ($r=0.49$), and F_w versus R ($r=0.48$).

The correlations calculated in Table IV were linear, but there is no reason to suppose that the functional dependencies between the variables should be linear. If the D_b and D_w versus d data are fit to linear, quadratic and cubic functions, the r values for the black pixel data are 0.54, 0.71, and 0.72 while the r values for the white pixel data are 0.48, 0.65, and 0.66, respectively. The fact that the r values do not increase between the quadratic and cubic fits suggests that there is a quadratic relationship between fractal dimension and cell differentiation, since the r values for quadratic fits for both the black and white pixel data are greater than 0.6. The black pixel quadratic fit reaches a minimum at a value of differentiation equal to 5.9 while the white pixel data fit reaches a maximum at a value of differentiation equal to 5.6. Since the the black pixel data comes closest to a fractal dimension value of 1.5 at its minimum, it suggests that the physical structure of progressive rat hepatoma becomes maximally complex at a value of differentiation between 5 and 6.

In order to examine the hypothesis put forward previously that maximum complexity occurs when an extremum occurs in the rate of change of the irreducible amount of information required to describe a system with respect to some parameter, the rate of change of I_{min} with respect to d was calculated. This is shown in Fig. 18. As can be seen, an extremum occurs at a value of $d=4.9$, which is close to the value of d suggesting maximum complexity from the fractal dimension data. An additional calculation assuming that I_{min} could be represented by maximum jpeg compression was also performed. The results were found to be essentially the same as those obtained using gzip compression with the extremum occurring at the same value of cell differentiation, $d=4.9$. The limited number of data points in this study does not allow us to claim strong support for the hypothesis, but the data is not inconsistent with it.

Analysis of the data suggests that: (1) consideration of the entire BCA curve can provide useful information above and beyond allowing for a single measure of fractal dimension, i.e., scales of interest where differences in image morphology occur, (2) there was no indication that the complimentary data set might have more correlation with disease state than the primary data set, (3) the relationship between disease state and measures of complexity is likely to be nonlinear, (4) there are indications that the this complex disease system passes through a state of maximum complexity as it progresses, and, finally, (5) the data is not inconsistent with the hypothesis that maximum complexity occurs when an extremum occurs in the rate of change of the irreducible amount of information required to describe a system with respect to some parameter. Thus the data provides some support for the idea that measures of complexity could be important elements of any future cancer "disease time" vector.

From a practical standpoint, variations in staining intensity and types of stains used would be expected to affect the processing and interpretation of images, particularly in choosing the threshold value for the conversion from grayscale to black and white images. For example, more intense nuclear staining (i.e., longer exposure to hematoxylin processing reagents) might cause smaller segments of nuclei (tangential cuts or partial nuclei) to be displayed and therefore increase the number of discrete items present for analysis. Likewise, the use of other stains that highlight other cell features (unique proteins or organelles) could provide additional information for analysis. In fact, one approach to describing tumor growth characteristics with morphologic measures is to use multiple histochemical and immunohistochemical stains on the same tumor sections as a means of assessing tumor pleomorphism, differentiation, and replicative state.

There is, however, clearly a need to standardize staining techniques to insure uniformity of image analysis between tumors. In this study, all tumors were cut to the same thickness, batch stained on an automated stainer at the same time, and examined on the same optical/image capture equipment. These standard procedures minimized variation, at least in this data set.

We have noted, during measurement, that the analytic methods we have used are highly suited to quantifying variability in different areas of the same tumor biopsy. Human pathologists attempt to perform an evaluation of cytologic variation during biopsy evaluation, but clearly do not quantify absolute variability nor ascertain if variation is key to prognosis. For example, it may be very important to determine the organization and cellular characteristics of the tumor margin cells, since they are most likely to represent the invasive interface with normal tissue. We are actively investigating the relationship between tumor behavior and variability of fractal measures within individual biopsies.

Perhaps the most significant aspect of this work is that there is a potential to develop a coherent suite of mathematical tools for defining benign tumor characteristics and malignant tumor characteristics. Not only might this analytical process quantify "typical" morphologic measures (nuclear size, shape and so forth) but could also provide a way to mathematically evaluate tumor organization. An added benefit could be the application of a computational intensive environment to augment tumor analysis without inherent pathologist bias created by experience and affected by both knowledge base and human ergometric factors.

This study was carried out on a very limited number of images and for that reason, the analysis is suggestive, not

definitive and a much larger study is indicated to test the implications of this preliminary analysis. The images processed for analysis in the research presented in this paper were in the standard format used by the pathologists in our laboratory. Although high resolution information is lost when jpeg compression is utilized, we felt that there was sufficient accurate data at the scales of interest to pathologists to allow our calculations to be meaningful. We appreciate that future studies should include parallel analysis of jpeg images with the same images in TIFF or other formats that do not degrade the high resolution data originally present in the images in order to test whether lossless image compression would produce better results.

We are currently collaborating with the Comprehensive Cancer Center of Wake Forest University to analyze of images of glioblastoma multiforme, images of canine malignant lymphoma and perform a comparison between fractal characteristics and WHO classification of lymphoma morphology. This study is ongoing and its results will be reported in the future. We will extend the present analytical techniques to include filtering of the images based on the color of stain used as well as "tuning" the threshold for conversion of grayscale to black and white images through critical value percolation analysis.

[1] R. Sole and B. Goodwin, *Signs of Life: How Complexity Pervades Biology* (Basic, New York, 2000).

[2] W.B. Spillman, Jr. *et al.*, Smart Mater. Struct. **5**, 247 (1996).

[3] *Proceedings of the 2001 Complex Adaptive Structures Conference*, edited by W.B. Spillman, Jr., Proc. SPIE Vol. 4512 (SPIE Opt. Eng. Press, Bellingham, 2001).

[4] W.B. Spillman, Jr., in *Proceedings of the 2002 European Workshop on Smart Structures in Engineering and Technology*, Proc. SPIE Vol. 4763 (SPIE Opt. Eng. Press, Bellingham, 2000), pp. 83–88.

[5] A.T. Winfree, *The Geometry of Biological Time*, 2nd ed. (Springer, New York, 2000).

[6] J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, 3rd ed. (Springer-Verlag, New York, 2003), p. 539.

[7] Y. Bar-Yam, *Dynamics of Complex Systems* (Addison-Wesley, Reading, MA, 1997).

[8] J.L. Connolly *et al.*, *Cancer*, 5th ed. (B.C. Decker, London, 2000), pp. 386–392.

[9] R. Blackwood and J.L. Robertson, "Effect of BRUCELLA RB51 adjuvant pretreatment on the growth of Morris 7777 induced tumors in Lewis rats" (in preparation).

[10] R.S. Cotran *et al.*, *Robbins Pathologic Basis of Disease*, 6th ed. (W.B. Sanders, Philadelphia, 1999).

[11] B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1982).

[12] C. Adami, *Introduction to Artificial Life* (Springer-Verlag, New York, 1998), pp. 175–198.

[13] P. Soille, *Morphological Image Analysis*, 2nd ed. (Springer-Verlag, New York, 2002).

[14] W.B. Spillman, Jr. and D.R. Huston, Opt. Eng. **37**, 898 (1998).

[15] G.A. Losa, Pathologica **87**, 310 (1995).

[16] S.S. Cross, J. Pathol. **182**, 1 (1997).

[17] D.S. Coffey, Nat. Med. **4**, 882 (1998).

[18] Y. Gazit *et al.*, Microcirculation (Philadelphia) **4**, 395 (1997).

[19] J.W. Baish and R.K Jain, Cancer Res. **60**, 3683 (2000).

[20] A.S. Kerenji *et al.*, Arch. Oncol. **8**, 47 (2000).

[21] S. Ghorui *et al.*, Pramana, J. Phys. **54**, L331 (2000).

[22] W.B. Spillman, Jr. *et al.*, Smart Mater. Struct. **11**, 623 (2002).

[23] H.D. Young, *Statistical Treatment of Experimental Data* (McGraw-Hill, New York, 1962).