Model for breast cancer diversity and spatial heterogeneity

J. Roberto Romero-Arias,[1,*] Guillermo Ramírez-Santiago,[2] Jorge X. Velasco-Hernández,[2] Laurel Ohm,[3] and Maribel Hernández-Rosales[4]

[1]*Conacyt. Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán 58040, Mexico*
[2]*Instituto de Matemáticas, Universidad Nacional Autónoma de México, Juriquilla Querétaro 76230, Mexico*
[3]*School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[4]*Conacyt. Instituto de Matemáticas, Universidad Nacional Autónoma de México, Juriquilla Querétaro 76230, Mexico*

(Received 28 February 2018; revised manuscript received 11 June 2018; published 4 September 2018)

We present and analyze a growth model of an avascular tumor that considers the basic biological principles of proliferation, motility, death, and genetic mutations of the cell. From an analysis of genomic data and considering the results of a regulatory network analysis we identify two sets of genes—a set of sixteen and six genes—that are believed to play an important role in the evolution of breast cancer. Considering that cancer cells shape the tissue microenvironment and niches to their competitive advantage, the model assumes that cancer and normal cells compete for essential nutrients and that the rate of mutations is affected by nutrients availability. To this end, we propose a coupling between the transport of nutrients and gene mutations dynamics. Gene mutation dynamics are modeled as a Yule-Furry Markovian process, while transport of nutrients is described with a system of reaction-diffusion equations. For each representative tumor we calculate its diversity, represented by the Shannon index, and its spatial heterogeneity, measured by its fractal dimension. These quantities are important in the clinical diagnosis of tumor malignancy. A tumor malignancy diagram, obtained by plotting diversity versus fractal dimension, is calculated for different values of a parameter β, that modulate proliferation rate. It is found that, when $\beta < 1$, tumors show greater diversity and more spatial heterogeneity as compared with $\beta > 1$. More importantly, it is found that the results and conclusions are similar when we use the six-gene set versus the sixteen-gene set.

DOI: 10.1103/PhysRevE.98.032401

I. INTRODUCTION

To date there is no consensus over how cancer is initiated; however, it is known that tumor growth occurs in several stages. The accepted general view is that a cell must undergo several gene mutations before it becomes cancerous. There are two kinds of mutations: "passenger" and "driver." The former are gene changes that do not affect cell fitness or contribute to cancer development, and they may appear and eventually vanish during any stage of tissue development. The latter are gene changes that are causally involved in cancer development, typically conferring a functional change as well as a somatic evolutionary advantage and play a crucial role in cancer progression [1,2]. Thus cancer development is the result of the gradual accumulation of driver mutations that enhance cell proliferation rate and inhibit cell death rate [3–5]. The detailed knowledge behind these mutations is unknown. Nonetheless, there is a general agreement that environment and heredity play important roles in cancer initiation. Tumor progression mainly involves oncogenes and tumor suppressor genes [6–9]. Oncogenes encode proteins that control cell proliferation and apoptosis [10], and can be activated by structural alterations resulting from mutation or gene fusion [11], by juxtaposition to enhancer elements [12], or by amplification. Translocations and mutations can occur as initiating events [13] or during tumor progression, while amplification usually occurs during progression. Activation of oncogenes by chromosomal rearrangements, mutations, and gene amplification confers a growth advantage or increased survival to cells carrying such alterations. All three mechanisms cause an alteration in the oncogene structure either by an increase or a deregulation of its expression [14]. On the other hand, tumor suppressor genes normally prevent unrestrained cellular growth, promote DNA repair, participate in the cell cycle checkpoint activation, and maintain the activity of every cell. In most cancers the "bad mutations" of tumor suppressor genes reduce functions and make cells grow without control [6,8]. In normal cells, however, hundreds of genes intricately control the processes of division and death, so that growth is the result of a balance between the activity of those genes that promote cell proliferation and those that suppress it. Cancer cells originate within tissues and no longer respond to many of the signals that control cellular growth and death. Over time, these cells become increasingly resistant to the molecular controls that maintain normal cells and, as a result, they divide more rapidly than their progenitors and become less dependent on signals from other cells. Cancer cells even evade programed cell death, despite the fact that their multiple abnormalities would normally make them prime targets for apoptosis. Phenotypic and functional heterogeneity usually arise among cancer cells within a tumor as a consequence of genetic variations, environmental differences, and irreversible changes in cellular properties. Cancer cell heterogeneity displays striking morphological, genetic,

*Corresponding author: romero@ifm.umich.mx

and proteomic variability and represents a great challenge to diagnosis, treatment, and drug resistance. Any tumor shows spatial heterogeneity that manifest as a nonuniform distribution of genetically distinct cancer-cell clusters. Analogously, temporal heterogeneity refers to the temporal variations in the genetic makeup of cancer cells [15]. We adopt this terminology for the rest of the paper.

It is known that spatial variations in cell genetic profiles lead to altered microenvironments which are observable through analysis of tissue pathology images [16]. Recent studies have demonstrated that heterogeneity is observed to varying extent across a wide variety of cancers, with the identification of both clonal and subclonal mutations [17–19]. Thus there are unlimited numbers of genetic and epigenetic alternatives along with environmental stress that contribute to tumor evolution. As a matter of fact, tumors reprogram pathways of nutrient acquisition and metabolism to meet the bioenergetic, biosynthetic, and redox demands of malignant cells. They also include symbiotic nutrient sharing, nutrient competition, and the role of metabolites as signaling molecules. These reprogrammed activities improve cellular fitness to provide a selective advantage during tumorigenesis, which leads to changes in signal transduction, the epigenome, and gene expression [20–22]. Therefore, a complex interplay of gene expression, DNA alterations, gene mutations, and environmental conditions are believed to be the main factors that drive tumor heterogeneity [23]. Hence the identification of a universal molecular mechanism at the center of cancer initiation and development is extremely difficult.

In particular, it has been shown that exposure of cultured cells to hypoxic conditions produces an elevated mutation frequency and a mutation pattern similar to those observed tumor-grown cells [24]. These findings suggest that the type of genetic instability found in malignant tumors may in part be the consequence of specific mutagenic properties of the hypoxic microenvironment. In addition, cancer cells also exhibit elevated glucose metabolism with increased glycolytic activity, so that fast growing tumors with poor or no vascularization are subjected to glucose starvation stress [25]. Thus the avascular tumor microenvironment can be considered hostile since it is characterized by areas of chronic or transient hypoxia, low pH, nutrient deprivation, and energy depletion [26–28]. Taking into account all these observations one simple way of incorporating some aspects of the tumor microenvironment is by considering the role of nutrients spatial gradients on cancer cell metabolism that influences the acquisition of new mutations. To this end, here we analyze a model based on the assumption that the probability of acquiring new mutations is related to the gradients of nutrients concentration.

There is a huge amount of models aimed at explaining different aspects of the complex growth of a tumor [29–33]. Most of them consider the spatiotemporal evolution of the tumor without incorporating the genetic information that drives the dynamics of growth [34–38]. This is an important aspect to be considered since most cancers are highly heterogeneous as a result of the mutations acquired during cell division. A three-dimensional model that combines the aspects of spatial dispersal and genetic evolution has been proposed. It describes the growth of primary tumors and metastases, as well as the development of resistance to therapeutic agents [39]. The model not only yields insight into spatial and temporal aspects of tumor growth, but also suggests that targeting short-range cellular migratory activity can have marked effects on tumor growth rates. It also shows how short-range dispersal and cell turnover can account for rapid cell mixing inside the tumor. Gene dynamics demonstrates that even a small selective advantage of a single cell within a tumor allows the descendants of that cell to replace the precursor mass in a clinically relevant time frame.

In the present paper we analyze a quantitative model of growth of an avascular tumor that considers cell proliferation, motility, death, and genetic mutations and the microenvironment through the quantification of the nutrient spatial gradients that are believed to affect the acquisition of new mutations [20–22,26–28]. We start from a reaction-difussion set of equations that describes the transport of essential and nonessential nutrients. These equations are coupled to a stochastic gene mutation dynamics of each cell modeled as a Markovian process. From genomic data, we identify two sets of six and sixteen genes, respectively, that are believed to play an important role in breast cancer tumor growth. Considering that quantitative measurements of diversity and spatial heterogeneity are important clues for clinical diagnosis of tumor malignancy, we estimate these properties by means of the Shannon diversity index and the fractal dimension. With these quantities we establish a tumor malignant-benign diagram for different values of a certain parameter β. It is found that for $\beta < 1$ the tumors display high genetic diversity and are spatially heterogeneous, whereas for $\beta > 1$ tumors develop less genetic diversity and are spatially less heterogeneous. The paper's layout is as follows: in Sec. II we present the model with the set of reaction diffusion equations for the transport of nutrients as well as the equations for stochastic mutation dynamics. In Sec. III we briefly explain the algorithms used to simulate the genes stochastic dynamics and the integration of the reaction-diffusion equations. In Sec. IV the results of the numerical simulations of tumors and the analysis of the diversity and spatial heterogeneity is presented. Finally, in Sec. V we discuss the results of the structure, diversity, and heterogeneity of tumors as well as the possible applications of the present quantitative modeling that may help in cancer diagnosis.

II. MODEL

Let us start with the reaction-diffusion model for the growth of an avascular tumor proposed by Ferreira *et al.* [40]. Tissue is made of three types of cells, namely normal, cancer, and necrotic cells, that live on a square lattice. Processes of proliferation, death, and competition for nutrients of normal and cancer cells are considered. It is assumed that essential and nonessential nutrients diffuse from the capillary vessel towards each cell throughout the tissue. Essential nutrients are glucose, aminoacids, vitamins, and minerals, and nonessential nutrients are oxygen, cholesterol, and vitamins that are made naturally in the body. These nutrients are critical for DNA synthesis and for cell proliferation; therefore, they are considered important for the development of gene diversity [41–43]. Accordingly, it is reasonable to assume that in active proliferating cells, mainly glucose is consumed and a significant portion of carbon is converted to lactic acid that is transported out of cells [44,45].

We assume that nutrient gradients increase mutation rate, while absence of or negligible nutrient gradients lead to a cell latent state decreasing the mutation rate. This hypothesis, introduced as a stochastic term coupled to the reaction-diffusion Eqs. (1) and (2), led us to show that the occurrence of mutations during tumor growth eventually yields high genetic diversity and tumor spatial heterogeneity, the hallmarks of most cancers.

The reaction-diffusion equations that describe the transport of essential and nonessential nutrients for cell proliferation are [40]

$$\frac{\partial N}{\partial t} = \nabla^2 N - \alpha^2 N\{\sigma_n + (1+\beta P_{dm})\lambda_N \sigma_c\}, \quad (1)$$

$$\frac{\partial M}{\partial t} = \nabla^2 M - \alpha^2 M\{\sigma_n + (1+\beta P_{dm})\lambda_M \sigma_c\}, \quad (2)$$

where $N(\vec{r},t)$ and $M(\vec{r},t)$ are the concentrations of essential and nonessential nutrients, respectively. The parameter α^2 represents the nutrient consumption rate for normal cells. Normal and cancer cells are represented by the field variables, $\sigma_n = 0, 1$ and $\sigma_c = 1, 2, 3, \ldots$, respectively. The ability of cancer cells to outcompete normal cells for essential and nonessential nutrients is denoted by λ_N and λ_M; usually they are greater than one so that cancer cells consume nutrients at a higher rate than normal cells [40]. The transport of nutrients on the right side of Eqs. (1) and (2) includes a term that is the product of the parameter β related to the cancer cell proliferation rate and the probability $P_{dm}(\vec{r},t)$ related to the occurrence of mutations at position $\vec{r}$ at time t. Hence the product βP_{dm} quantifies the accumulation of mutations in each cancer cell in the proliferation. This product may not always increase, but could also remain constant or decrease during tumor progression. This stochastic term modifies the reaction term that accounts for the ability of cancer cells to compete for nutrients. Note that when $\beta = 0$, Eqs. (1) and (2) reduce to the nutrient transport equations in [40]. In order to obtain tumor structures that present a compact morphology with a few branches we chose the parameter values $\lambda_M = 10$, $\lambda_N = \{50, 100\}$, and $\alpha = \{4\times10^{-3},\ 6\times10^{-3}\}$ [40]. We have set $\lambda_N > \lambda_M$, so that the consumption of essential nutrients is greater than that of nonessential nutrients. These transport equations are complemented with the probability of death. This probability can be expressed as [40]

$$P_{\text{death}}(\vec{r},t) = \exp\left[-\left(\frac{M(\vec{r},t)}{\theta_{\text{death}}\sigma_c(\vec{r},t)}\right)^2\right], \quad (3)$$

which represents the probability of death of a cancer cell located at point $\vec{r}$ at time t, described by the field variable $\sigma_c(\vec{r},t) = 1$, which means that there is one cancer cell in the tissue. The concentration of nonessential nutrients—for cell proliferation—at point $\vec{r}$ at time t is denoted by $M(\vec{r},t)$, while θ_{death} controls the width of the curve. Considering the case of avascular tumors with small necrotic cores [31,46,47], we chose $\theta_{\text{death}} = 0.01$. This value reflects the fact that the probability of death increases when the availability of oxygen is about 2%. As the value of θ_{death} increases the size of the necrotic core increases too [40].

During cancer cell proliferation (A) there occur two independent processes or events: (i) random mutations (A_u) and (ii) nutrients consumption (A_v). Then, the probability of cancer cell division at point $\vec{r}$ at time t can be written as

$$P_{\text{div}}(A) = P(A|A_u)P(A_u) + P(A|A_v)P(A_v), \quad (4)$$

where $P(A|A_u)$ is the probability of division given that random mutations occur and $P(A|A_v)$ is the probability of division given nutrient consumption. In Eq. (4), $P(A_u)$ is the probability that a random mutation occurs, while $P(A_v)$ is the nutrient consumption probability by cancer cell. We do not need the explicit expressions for these probabilities as will be seen below. To establish a correlation between the occurrence of random mutations given that a cell division happened, we use Bayes' property to rewrite the first term on the right-hand side (RHS) of Eq. (4). That is, $P(A|A_u)P(A_u) = P(A_u|A)P_{\text{div}}(A)$, so that Eq. (4) can be recast as

$$P_{\text{div}}(A) = P(A_u|A)P_{\text{div}}(A) + P(A|A_v)P(A_v). \quad (5)$$

Solving for $P_{\text{div}}(A)$ we obtain the probability of cancer cell division:

$$P_{\text{div}}(A) = \frac{P(A|A_v)P(A_v)}{1 - P(A_u|A)}. \quad (6)$$

We now assume that the numerator of this equation favors mutations through nutrients concentration as [40]

$$P(A|A_v)P(A_v) = \left(1 - \exp\left[-\left(\frac{N(\vec{r},t)}{\theta_{\text{div}}\sigma_c(\vec{r},t)}\right)^2\right]\right), \quad (7)$$

where θ_{div} controls the shape of this probability. We chose $\theta_{\text{div}} = 0.3$ in order to obtain a compact tumor structure since as θ_{div} increases the tumor becomes more branched [40]. On the other hand, we assume that the probability of random mutations given a division processes occurs is $P(A_u|A) \equiv \beta P_{\text{dm}}$; then the probability of cancer cell division can be expressed as

$$P_{\text{div}}(\vec{r},t) = \frac{1 - \exp\left[-\left(\frac{N(\vec{r},t)}{\theta_{\text{div}}\sigma_c(\vec{r},t)}\right)^2\right]}{1 - \beta P_{\text{dm}}(\vec{r},t)}. \quad (8)$$

This equation represents the probability of division of a cell located at point $\vec{r}$ at time t. The concentration of essential nutrients at point $\vec{r}$ at time t is represented by $N(\vec{r},t)$. Since $P_{\text{div}} \leqslant 1$, then $\beta \leqslant \exp[-(\frac{N}{\theta_{\text{div}}\sigma_c})^2]/P_{\text{dm}}$; that is, proliferation and mutation rates depend on nutrients concentration. In this sense, nutrients play the role of a *"catalyst"* for mutations. Note that when the nutrient concentration is large the right side of this inequality is small whereas when nutrient concentration is small the right side of this inequality is large. Therefore, regions with high nutrient concentration favor cell survival and increase mutation rate, and regions with low nutrient concentration disfavor mutations.

A. Gene types

To determine the genes to be incorporated into the genetic dynamics, we analyzed data from the Data Release 23 from the European Union breast cancer project (BRCA-EU). We found a set of sixteen genes that are believed to play a major role in breast cancer development. The number of mutations of each of these genes are shown in Fig. 1. In this set we can identify the tumor suppressor genes: TP53, ATR, ATM, E2F1; oncogenes: BRCA1, ERBB2, MDM2, NRAS, HRAF; and kinase regulators: CHEK2, KRAS, CHECK1, AKT1, CDK2.

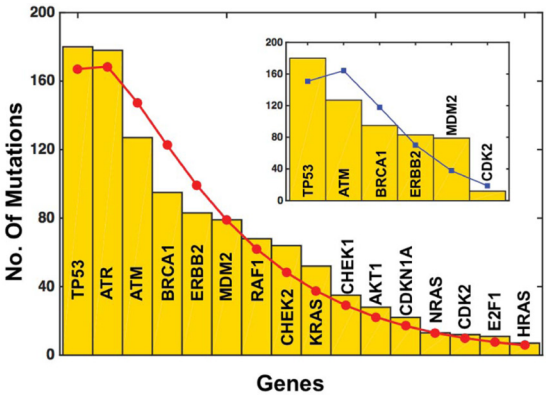

FIG. 1. Histogram of the number of mutations of the sixteen-gene set that are believed to play a major role in breast cancer development. Genes have been ordered according to their number of mutations. Data is taken from Breast Cancer ICGC Project (https://dcc.icgc.org/) for the European Union. The inset shows the number of mutations of the six-gene set that plays a crucial role in breast cancer as suggested from a regulatory network analysis [48]. The red line with circles and the blue line with squares represent the results of fitting a negative binomial distribution to each histogram. See text for more details.

On the other hand, a recent regulatory network analysis in breast cancer [48] inferred from gene expression data suggests that there are six genes, namely, TP53, ATM, ERBB2, BRCA1, MDM2, and CDK2, that play a major role in breast tumor development. This finding appears to be consistent with the general belief that driver mutations are produced mainly by six genetic variations [5,6,8,9,42,48,49]. Based on this data we focus our attention on two histograms that represent the number of mutations of each gene. The histogram shown in the main figure—Fig. 1—is related to the sixteen genes set, while the one shown in the inset of Fig. 1 is related to the six genes set [48]. We emphasize that gene dynamics depends only on the statistical distribution of mutations, that is, the number of mutations of each gene during cancer evolution. Therefore, our model of the genetic dynamics is independent of the detailed biological properties of the genes and there is no need to identify the driver mutations in the data set [50].

Mutations can be modeled by a negative binomial distribution with parameters p and r [51–57]. Thus we fitted these distributions to the data represented by the histograms and the results are shown with solid lines in Fig. 1. The sixteen-gene fitting shown in Fig. 1 with a red line and circles yielded the results $p = 0.2578 \pm 0.0136$ and $r = 1.3591 \pm 0.0877$ with a goodness of fit $Q = 0.9301$ and correlation coefficient $R = 0.9779$. Similarly, the result for the six-gene fitting is plotted in the inset of Fig. 1 with a blue line and squares. This fitting yielded the results $p = 0.6560 \pm 0.0481$ and $r = 3.1189 \pm 0.6529$ with a goodness of fit $Q = 0.8713$ and correlation coefficient $R = 0.8699$. The results of this statistical analysis are used in the implementation of the stochastic simulations of tumor growth in Sec. III A.

Since our analysis suggests that the results obtained with both sets are consistent and robust, it is sufficient to consider the six-gene set. In Sec. IV we elaborate more about these findings.

B. Mutation dynamics

By assuming that cell mutations depend only on the previous genetic state, one can model the mutations dynamics by means of a Yule-Furry process [58]. The corresponding master equation is

$$\frac{dP_x(\vec{r},t)}{dt} = -\gamma(\vec{r},t)xP_x(\vec{r},t) + \gamma(\vec{r},t)(x-1)P_{x-1}(\vec{r},t), \tag{9}$$

where $P_x(\vec{r},t)$ represents the probability that the cell at position $\vec{r}$ at time t in the tissue undergoes x ($x = 0, 1, 2, \ldots$) mutations and $\gamma(\vec{r},t) > 0$ is the jump probability that one new mutation, $x \to x+1$, will happen in the time interval $[t, t+dt)$. The solution of Eq. (9) is the geometric distribution, $P_x(\vec{r},t) = (1-p(\vec{r},t))^{x-1}p(\vec{r},t)$, with $p(\vec{r},t) = \exp(\int_0^t \gamma(\vec{r},\tau)d\tau)$ [58]. In what follows we relate this jump probability to the microenvironment. As argued in the Introduction, nutrients spatial gradients modify the mutation rate [24–26]. One simple way of incorporating this hypothesis into the mutation dynamics is to assume that the jump probability of new mutations depends on these spatial gradients. That is, we make the *anzatz*, $p(\vec{r},t) = \exp(\int_0^t \gamma(\vec{r},\tau)d\tau) = \exp[-(\frac{N(\vec{r},t)}{\theta_{\rm div}})^2]$, where $N(\vec{r},t)$ represents the concentration of essential nutrients at position $\vec{r}$ at time t in the tissue. This approach is consistent with recent findings in cancer development that suggest that interactions between cancer cells and their tissue habitat are reciprocal, that is, cancer cells shape the tissue microenvironment and niche to their competitive advantage [18,59].

It is known that driver mutations can be randomly activated by structural alterations resulting from mutation or gene fusion, by juxtaposition to enhancer elements, or by amplification of random mutations acquisition, that can be modeled by a Poisson process [6,41,60–62]. Then, one can write the total probability distribution of having mutations at a cell located at position $\vec{r}$ in the tissue at time t as the product of two independent probabilities:

$$P_{\rm dm,j}(\vec{r},t) = G(p(\vec{r},t),z_j)N_\lambda(\lambda,k_j), \tag{10}$$

where $G(p(\vec{r},t),z_j)$ is a geometric probability distribution with mean $(1-p(\vec{r},t))/p(\vec{r},t)$. Here $p(\vec{r},t)$ is given by the *anzatz* indicated above and z_j is the number of viable mutations of gene j. The function N_λ is a Poisson probability distribution with mean λ and represents the probability of occurrence of k_j mutations of gene j per cell. This factorization will play an important role in the implementation of the stochastic simulations for the tumor gene dynamics.

Now let us estimate the upper bound for the parameter β in Eqs. (1) and (2). Recall that $\beta < p/P_{\rm dm}$; since the mean value of Eq. (10) is $(1-p)\lambda/p$, one can obtain the following upper bound for β:

$$\beta < \frac{p^2}{\lambda(1-p)}. \tag{11}$$

Since the jump probability of new mutations is a Gaussian function of $N(\vec{r},t)$ and because there occur few random mutations we can assume $\lambda \sim O(10^1)$ [26]. Then one finds that, for $N(\vec{r},t) \sim 1$, the jump probability is $p \sim O(10^{-1})$ and, on the

contrary, when $N(\vec{r}, t) \sim 0$ one gets $p < O(10^{-5})$. Therefore, the jump probability satisfies the inequality $O(10^{-5}) \leqslant p \leqslant O(10^{-1})$. From these inequalities we conclude that $0<\beta<10$. These bounds for β help us to explore two limiting regimes for mutations during proliferation: (i) low mutation rate, when the cell is in a *"latent state"* due to low nutrients availability, and (ii) high mutation rate when there are plenty of nutrients. Our model postulates that there is an intrinsic nonlinear coupling between the master equation for mutation dynamics and the reaction-diffusion system that describes the nutrient concentration at each cell in the tissue. To fully describe the dynamics of this complex system we need to perform stochastic simulations as described in the forthcoming section.

III. NUMERICAL SIMULATIONS

A complete flow diagram for the model simulations of the tumor evolution is shown in Fig. 2. This diagram shows in a synthetic form the spatiotemporal evolution of the tumor including the gene mutations dynamics. Let us begin with a brief explanation of the algorithm that was used to simulate the genes stochastic dynamics.

A. Stochastic simulations

To simulate the stochastic mutation dynamics we used the Tau-Leaping Gillespie algorithm [63], which has demonstrated its usefulness in the simulations of different processes in molecular biology. Let us assume that, at a given time t, the state of the system is defined by the vector $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$, in which each coordinate represents the number of mutations in each of the n genes. Then, the change of this state vector in the time interval $[t, t+\tau)$ is given as

$$\mathbf{x}(t+\tau) \rightarrow \mathbf{x}(t) + \sum_j \mathbf{k}_j \nu_j, \tag{12}$$

where $\mathbf{k}_j$ is a vector of random numbers generated from a Poisson distribution with mean $\mathbf{a}_j(\mathbf{x})\tau$ and ν_j is the vector that changes the mutations of gene j, by zero, or 1. The selection of gene j is made by choosing a random number from a set of numbers distributed according to the negative binomial distribution obtained from the fits to the histograms in Fig. 1.

Let $\mathbf{a_j(x)}$ be the propensity functions that represent the probability of having one mutation at time t of gene j in the set. Since mutations occur with equal probability regardless of the gene chosen the values of $\mathbf{a_j(x)}$ can be set equal to one for every gene. Therefore, the time τ required for the number of mutations to increase by one unit is

$$\tau = \frac{1}{\mathbf{a_0(x)}} \ln\left(\frac{1}{r_j}\right), \tag{13}$$

where $\mathbf{a_0(x)} = \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{a_j(x)}$ and r_j is a random number uniformly distributed in the interval [0,1] corresponding to gene j. Since the dynamics of the tumor evolution is the result of the coupling of gene mutations and nutrient dynamics, an extended version of the tau-leaping method was applied to obtain an effective sampling of the biophysical relevant quantities [64–67]. Thus the change of the system's state $\mathbf{x(t)}$ during a time τ occurs in accordance with the following

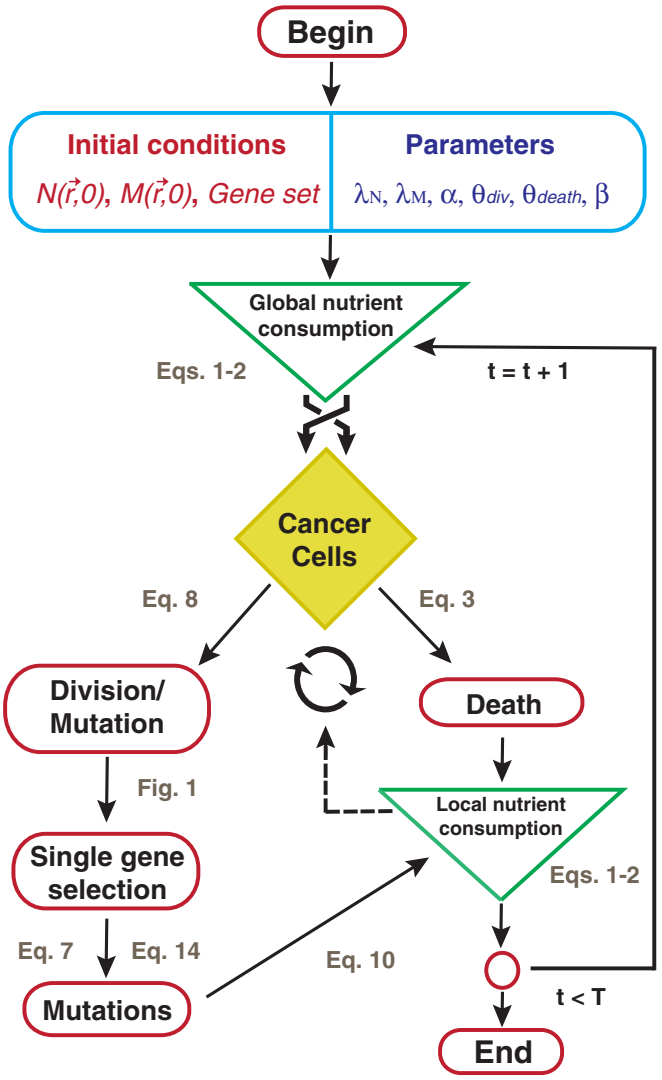

FIG. 2. Flow diagram for tumor evolution. The equations that are required at each simulation step are indicated by their number in the text. The relation with the genomic data (histograms shown in Fig. 1) is also indicated in between the division mutation and single gene selection steps. The letters T and t denote the cell generation time and the cell cycle, respectively. Cancer cells are chosen at random (indicated by the shuffle symbol) until a complete sweep over all cells is completed. The recycle symbol indicates that the cancer cells feed from local nutrient gradients.

equation:

$$\mathbf{x}(t+\tau) \rightarrow \mathbf{x(t)} + \sum_{\mathbf{j}} \mathbf{k_j z_j} \nu_{\mathbf{j}}, \tag{14}$$

where $\mathbf{z_j}$ is a vector formed with random numbers, which are distributed according to a geometric distribution, and $\mathbf{k_j}$ is the random vector generated from a Poisson distribution. We chose these random number distributions because mutations dynamics is described by the product of these two probability distributions. See Eq. (10).

B. Numerical integration

The solutions of the reaction-diffusion system, Eqs. (1) and (2), together with the probabilities for division and death,

Eq. (8) and Eq. (3), respectively, are calculated numerically. At the beginning of the simulations all cells are normal except for one that has developed cancer and is located near the lattice center. We assume that normal, cancer, and necrotic cells are on the sites of a square lattice of size $L \times L = 500 \times 500$ [40]. Essential and nonessential nutrients for cell proliferation are continuously supplied through a capillary located at the top of the lattice simulating the bloodstream. To avoid steep gradients in the nutrients concentrations, Eqs. (1) and (2) are integrated using a sublattice of size 10×10 units around each site populated with cancer cells. To this end we applied zero flow boundary conditions in all sides until the steady state was reached. After passing by all the sites populated with cancer cells the reaction-diffusion equations are solved globally with zero flow boundary conditions in all sides except on the top, where the bloodstream is located. This procedure completes one simulation cycle (see Fig. 2). From now on one simulation cycle will be denoted as a generation time T.

We assume that the initial cancer cell suffers mutations in the tumor suppressor gene TP53, which is the one that plays a crucial role in tumor growth. After a cell division occurs the daughter cell position is chosen randomly as one of the four nearest neighbors of the mother cell position. This descendant undergoes mutations according to the probability distribution given in Eq. (7). To decide the occurrence of mutations during the cell division process we choose a random number q distributed uniformly in the interval [0,1] and compare it with the probability given in Eq. (7). A mutation process occurs if $q > P(A|A_v)P(A_v)$; otherwise, it is rejected. A new cycle is initiated by choosing randomly a cancer cell with equal probability. The simulation cycle is completed by passing by all the lattice sites.

The time length of the simulations is estimated by tuning the number of mutations of each gene with that of the genomic data. See Fig. 1. This estimation led us to conclude that, on average, a simulation of 800 cycles is sufficient for each gene to reach the number of mutations of the genomic data. A typical simulation of this length yielded tumors with size smaller than 450×450 lattice sites for most combinations of values of the model parameters considered here. To understand the statistical meaning of the results we performed averages over 5, 10, and 20 simulations. It was found that the results were consistent within one standard deviation with those corresponding to just one simulation. Therefore, the results we report here correspond to one simulation of the system. We carried out simulations for $\beta = 0.0,\ 0.25,\ 0.5,\ 0.75,\ 1.0,\ 2.0$, and 4.0. These values were chosen in order to explore the two mutation rate limits described at the end of Sec. II B.

IV. RESULTS

This section presents and discusses the results obtained from the numerical simulations of the model explained in Sec. II. Let us begin by considering the set of six genes obtained from the regulatory network analysis. Figure 3 presents the tumor spatial distribution of mutations for each one of the six genes after one simulation cycle. The results have been arranged in a clockwise direction according to a decreasing number of mutations starting from the TP53 gene, which accumulates the most mutations. The central figure represents

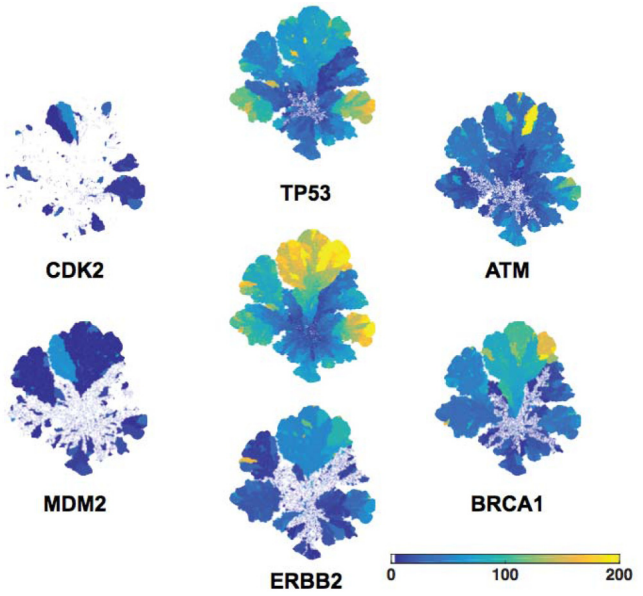

FIG. 3. Starting from the top cluster and going in the clockwise direction are shown the spatial distribution of mutations of the six genes set. They are ordered in a decreasing way according to the number of mutation of the genes (TP53, ATM, BRCA1, ERRB2, MDM2, and CDK2) indicated in the inset of Fig. 1. At the center lies the tumor showing the spatial distribution of mutations of the six genes altogether. The results correspond to the parameter values: $\alpha = 4 \times 10^{-3}, \lambda_N = 100$, and $\beta = 1$, after $T = 800$ cycles. The scale of colors is related to the number of mutations for each gene.

the superposition of the spatial distribution of mutations in the six genes under consideration. Note the accumulation of mutations at the upper tumor periphery. This is expected since the nutrient capillary supply is located at the top of the tissue domain and the nutrient concentration gradient drives both cell division and mutation. The results suggest that the tumor structure is defined by the major accumulation and spreading of mutations of the tumor suppressor genes TP53 and ATM. In addition, the central part populated with cancer cells (white region) represents the spreading of genes BRCA1, ERBB2, MDM2, and CDK2 that accumulate fewer mutations, suggesting that their contribution to the early stages of tumor progression is not relevant. However, genes CDK2 and MDM2 accumulate mutations in cells located in a smaller region close to the periphery. The observed spatial distribution of mutations in these figures suggests that tumor structure develops a spatial heterogeneity and certain degree of gene mutation diversity. It looks similar to what is observed in the clinical analysis of biopsies [68–71].

Tumor heterogeneity and diversity are quantified by the k-means clustering algorithm [72] and Shannon entropy index [73], respectively. The k-means clustering algorithm yields a graphical distribution of cell clusters according to the number of mutations and is referred to the initial position of the tumor center of mass [74]. The Shannon entropy is defined as $\mathcal{H} = -\sum_i P_i \ln P_i$, where P_i is the probability that i mutations occurred in the whole cancer tissue. To compute P_i we counted the number of cells that underwent one mutation, two mutations, three mutations, etc., and then we divided this quantity

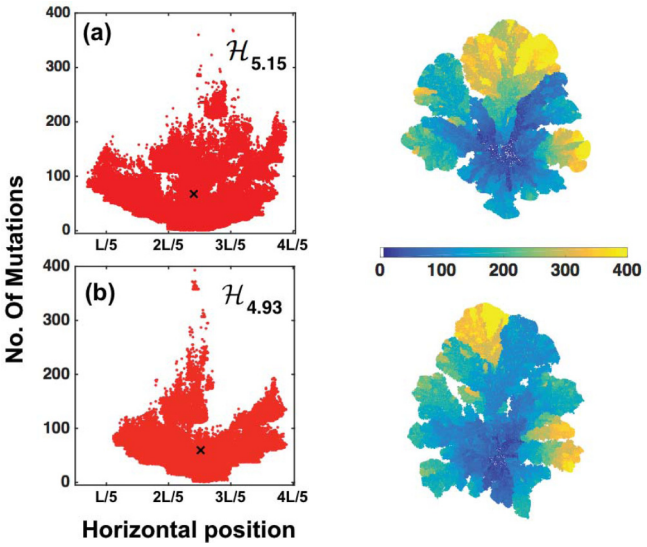

FIG. 4. Clusters obtained for the same model parameters and time as in Fig. 3. (a) For the six genes set we obtained a diversity $\mathcal{H} = 5.15$. (b) For the sixteen genes set we obtained a diversity $\mathcal{H} = 4.93$. These values correspond to a high diversity tumor. Note that cells with a high number of mutations are located in the tumor periphery, far from the tumor center of mass indicated with the symbol $\times$. The scale of colors is related to the number of mutations for each gene and $L = 500$ is the horizontal lattice size.

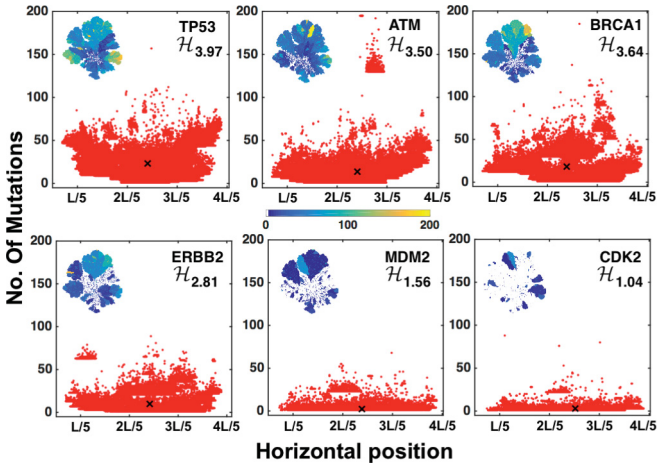

FIG. 5. Clusters indicate that large number of mutations (for genes TP53, ATM, BRCA1, ERRB2, MDM2, and CDK2) are located far from the tumor center of mass ($\times$) in the six-gene set. The inset in each figure represents the tumor spatial distribution of the corresponding gene. Observe that the diversity $\mathcal{H}$ decreases as the cluster size decreases. These results were obtained for the same parameter values as in Fig. 3. The color bar at the middle of the centered column measures the number of mutations and $L = 500$ is the horizontal lattice size.

by the total number of cancer cells. Therefore, the Shannon index in our case is a measure of the diversity of cell types characterized by the number of mutations. In ecology, Shannon index values lying in the range $1.5 < \mathcal{H} < 3.5$ are considered as a normal diversity of species [73]. However, $\mathcal{H} > 4$ indicates a very rich community. In the present case an increase in $\mathcal{H}$ is directly related to an increase of gene mutations due either to the existence of a large number of cells with a relatively small amount of mutations or a few cells with a large amount of mutations. The Shannon index can be measured clinically through immunohistochemistry staining to evaluate cell-level heterogeneity as well as patients' therapeutic response [75]. Recent clinical studies in breast cancer have found that genetic diversity fluctuates in the range $1 < \mathcal{H} < 4$ [76,77]. This range of fluctuations agrees reasonably well with the interval of $\mathcal{H}$ that is considered as normal diversity in ecology. Hence, from now on, we will consider the upper bound $\mathcal{H} = 3.5$ as transition point for diversity. That is, $\mathcal{H} < 3.5$ will indicate normal diversity, while $\mathcal{H} > 3.5$ will be referred to as high diversity.

In what follows we use the k-means clustering algorithm to obtain the graphical distribution of cells according to the number of mutations. We use the notation "horizontal position" to denote the abscissas in the clusters shown in Figs. 4–8. Figure 4 shows cell clusters formation of a tumor where the number of mutations of each cell is plotted against its position in each row of the 2D lattice. These results were obtained for the parameter values $\alpha = 4 \times 10^{-3}$, $\lambda_N = 100$, and $\beta = 1$, after completing 800 cycles of simulation. We choose this value of β because it represents a transition point in the accumulation of mutations during proliferation as will be seen below. In Fig. 4(a) are shown the cell clusters and the corresponding tumor obtained with the six-gene set. The tumor has $\mathcal{H} = 5.15$. Figure 4(b) shows the cell clusters and tumor obtained with the sixteen-gene set. It is clear that it has a lower diversity, $\mathcal{H} = 4.93$, as compared to the tumor in Fig. 4(a). This is expected since as the number of genes increases the probability that each gene undergoes a number of mutations decreases because there are more genes available in which mutations may happen. Observe that cell clusters that underwent the larger number of mutations are located at the tumor periphery, far from the tumor center of mass denoted by the symbol $\times$. The spatially heterogeneous structure and high genetic diversity of the tumor is due to the occurrence of mutations in different genes at different tumor positions. Since the values of diversity indices in both the sixteen and the six-gene sets are similar, we are confident that our results are robust.

In Fig. 5 are shown the clusters for the six-gene set after $T = 800$ simulation cycles. The corresponding diversity is also indicated. Observe that the cluster structure varies across the tumor because the number of mutations of each gene is different. Besides, diversity becomes greater as the number of mutations increases. Reciprocally, as the number of mutations decreases, so does the diversity. These results suggest that both the tumor suppressor gene TP53 and the oncogene BRCA1 are important in the structure of tumor growth since both have high diversity: $\mathcal{H} = 3.97$ and 3.64, respectively. This suggests that genes with diversity greater than 3.5 are important in breast cancer development which is consistent with the attractors found in the gene regulatory network analysis of [48]. Thus tumor genetic structure and diversity obtained from the present model are consistent with the gene regulatory network analysis. In addition, the ATM gene is considered to play a central role in the signal transduction of early stages of tumor progression and has $\mathcal{H} = 3.5$ as the transition point for diversity.

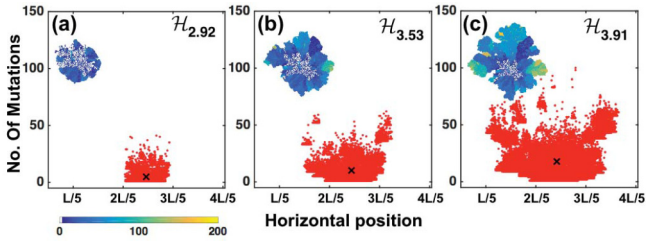

FIG. 6. Cluster evolution of gene TP53 at three stages of the tumor growth: (a) $T = 200$ cycles, (b) $T = 400$ cycles, and (c) $T = 600$ cycles. On the upper left side of each frame are shown the corresponding spatial distributions in the full tumor. The color bar indicates the number of mutations and $L = 500$ is the horizontal lattice size. In this case the diversity increases from 2.92 to 3.91. These results were obtained for model parameters as in Fig. 3.

The other two genes, ERBB2 and MDM2, known as oncogenes have a diversity that can be considered normal; however, the gene CDK2 has the lowest diversity, $\mathcal{H} = 1.50$.

Figure 6 presents the cluster evolution of the distribution of mutations of gene TP53 as well as at three stages of tumor growth. Cells that underwent the largest number of mutations are always related to clusters that are on the tumor periphery, while cells that suffered few mutations are related to clusters located around the tumor center of mass. The diversity of the first and second cluster configurations indicates a tumor with normal diversity, while the third configuration has a diversity greater than 3.5 indicating a high diversity. The same analysis was performed for the other five genes and the conclusions were similar.

Figure 7 illustrates the cluster structure, left and right columns, for the sixteen gene set after a simulation time of $T = 800$ cycles. The model parameter values used for these results are $\alpha = 4 \times 10^{-3}$, and $\lambda_N = 100$ for the four values of β: (a) $\beta = 0.25$, (b) $\beta = 0.75$, (c) $\beta = 2.0$, and (d) $\beta = 4.0$. We chose these values of β to understand the cluster and tumor structure. The corresponding tumors together with the color bar as a reference are plotted at the middle of the figure. This indicates that as β increases the tumor diversity decreases. These results point to a close relationship between the cluster aggregation and the branched structure of the tumor. Figure 7(a) shows the cluster structure for $\beta = 0.25$ with $\mathcal{H} = 4.99$, as an indication of high diversity. Observe that the largest number of mutations occurs right above the cluster's center of mass, indicated by $\times$. However, a large number of mutations is observed at both sides of the tumor center of mass. As a consequence, the tumor becomes branched, with the top region (yellow part) being the one with the highest number of mutations. The cluster shown in Fig. 7(b) corresponds to $\beta = 0.75$ with a diversity $\mathcal{H} = 5.0$. In this case the region with the largest number of mutations is positioned at the upper right side, again in the tumor periphery. The cluster shown in Fig. 7(c) corresponds to $\beta = 2.0$ and has a diversity $\mathcal{H} = 4.39$. There one sees that the number of mutations decreases by about half compared to the number of mutations accumulated in the clusters in Figs. 7(a) and 7(b). Notice that right above the cluster's center of mass there is a bifurcation of two clusters that represent two regions of the tumor that suffered a large number of mutations. In this case the tumor developed fewer branches

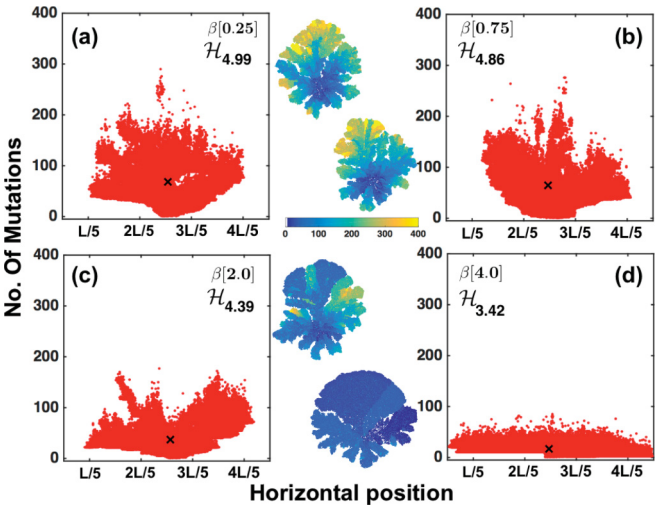

FIG. 7. Tumor (at the middle column) and cluster structure (left and right columns) after a simulation time $T = 800$ cycles for values of β: (a) $\beta = 0.25$, (b) $\beta = 0.75$, (c) $\beta = 2.0$, and (d) $\beta = 4.0$. As the parameter β increases the diversity $\mathcal{H}$ decreases. These results are obtained for $\alpha = 4 \times 10^{-3}$ and $\lambda_N = 100$. In this case the tumor shows a branched structure at the top of the periphery indicating the cells located there underwent a large number of mutations. The color bar represents the number of mutations and $L = 500$ is the horizontal lattice size.

and became more compact at the middle. Figure 7(d) shows the cluster obtained for $\beta = 4.0$, which has a diversity $\mathcal{H} = 3.42$. For this value of β the cluster is less branched, more compact, and homogenous, indicating that most of the tumor cells suffered about the same low number of mutations. Since tumor diversity is the smallest of the four, this suggests that most of the cells conserved their genetic lineage during proliferation. This conserved lineage can be observed in the tumor structure shown at the center of Fig. 7(d) where there are two main distributions of mutations—dark and light regions—that spread from the center up to the periphery.

Considering altogether the clusters shown in Figs. 4 and 7, one sees that the cluster that shows the highest diversity corresponds to $\beta = 1$. More importantly, from the analysis of Fig. 7, one finds that the larger the value of β, the lower the number of mutations, while the tumor structure becomes less branched and more compact. At this point it is important to bear in mind that in the present model the role of the nutrients in the tumor microenvironment is accounted for by considering the competition for essential nutrients during cell proliferation between cancer and normal cells. In fact, the nutrients transport equations, Eqs. (1) and (2), contain the parameters α and λ_N that measure, respectively, the nutrient consumption rate of normal cells and an additional factor, $\beta P_{\rm dm}$, by which nutrient consumption by cancer cells differs from their normal counterparts. Therefore, in our model we assumed that mutations leading to cancer confer higher fitness and thus a competitive advantage. Usually λ_N is chosen to be greater than 1 so that cancer cells consume essential nutrients at a higher rate than normal cells [40].

Figures 8(a)–8(d) show the cell clusters as well as the full tumor spatial distribution of mutations for four combinations

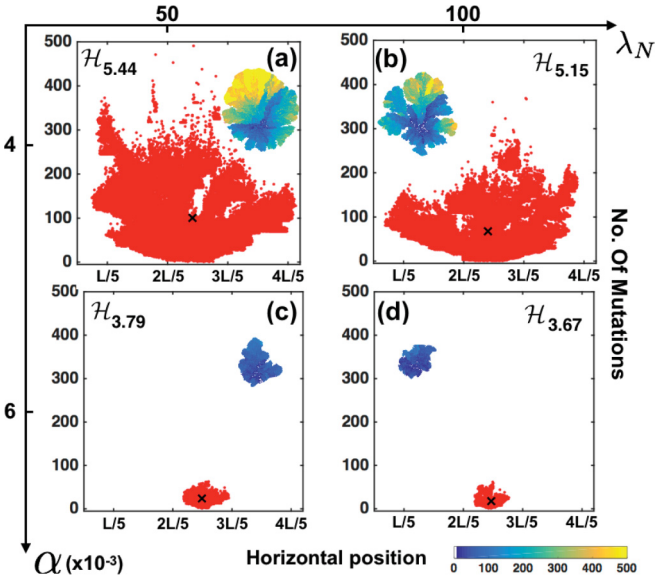

FIG. 8. Cluster structure and spatial distribution of mutations in the tumor for representative combinations of parameter values, α and λ_N, for $\beta = 1$, and simulation times $T = 800$ cycles. (a) $\alpha = 4 \times 10^{-3}$ and $\lambda_N = 50$, (b) $\alpha = 4 \times 10^{-3}$ and $\lambda_N = 100$, (c) $\alpha = 6 \times 10^{-3}$ and $\lambda_N = 50$, and (d) $\alpha = 6 \times 10^{-3}$ and $\lambda_N = 100$. The index of diversity values are also indicated for each case. These values decrease as the tumor becomes more compact and smaller in size. The color bar represents the number of mutations and $L = 500$ is the horizontal lattice size.

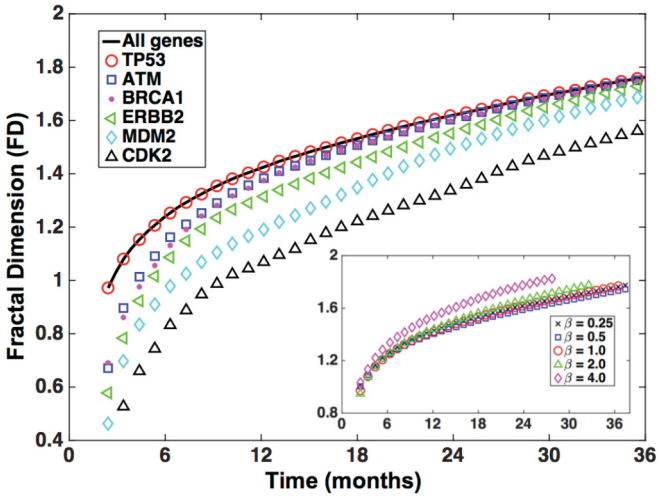

FIG. 9. FD time evolution for the spatial distribution of each gene for the six-gene set. The spatial distribution of mutations of gene TP53 shows a FD time evolution that is quantitatively similar to that of the whole tumor (solid black line). The results represent a tumor with parameters: $\alpha = 4 \times 10^{-3}$, $\lambda_N = 100$, and $\beta = 1$. The inset shows the time evolution of the whole tumor FD for different values of β and same values of α and λ_N. These results indicate that as the number of mutations increases FD approaches one universal curve.

of values of the parameters α and λ_N. The clusters and tumors have been referred to a coordinate system whose vertical axis represents the values of α, while the horizontal axis represents the values of λ_N. The parameter β has been assigned the value $\beta = 1$. It is found that, for $\alpha = 4 \times 10^{-3}$ and $\lambda_N = 50$, the tumor is compact and cells located at the tumor periphery undergo up to 500 mutations, as indicated by the cluster analysis in Fig. 8(a). However, in Fig. 8(b) the tumor develops fingerlike structures with a smaller main core than in Fig. 8(a). In addition, the cluster structure indicates that at the periphery (top right) there is a small cluster which is related to the tumor top branch, suggesting that between 200 and 400 mutations occurred. These results indicate that accumulation of mutations is directly related to cell fitness and diversity. In Figs. 8(c) and 8(d), $\alpha = 6 \times 10^{-3}$ and $\lambda_N = 50,\ 100$, respectively. The cluster and tumor structures become compact and decrease in size. In these cases the number of mutations that occur is a fraction (between 0.2 and 0.3) of the accumulation of mutations of the clusters in Figs. 8(a) and 8(b) and the diversity value becomes smaller than four, which indicates normal diversity. These results suggest that, for a given value of λ_N as α increases, the overall nutrients consumption rate—by normal and cancer cells—is higher and cancer cell fitness and diversity decreases.

Another quantity that characterizes the tumor structure is the fractal dimension (FD). This quantity can be measured in histopathology slides of tissue samples and is an important step in the diagnosis of it [78–80]. In addition, the change in texture or appearance of distortions in breast cancer tumors can be detected from mammograms by estimating the FD [81]. With the aim of relating these clinical measurements with the tumor structure and the spatial distribution of mutations of the *in silico* tumors—for instance those presented in Fig. 3—the time evolution of the FD was calculated by means of the standard box-counting algorithm (Fig. 9). FD increases monotonically as a function of time for each of the six genes and also becomes systematically greater for those genes that underwent more mutations, as expected. That is, when one or more genes suffer many mutations tumors become more diverse and heterogeneous. To express the FD time evolution in terms of a biological time scale one can relate one simulation cycle T with a biological cell division cycle that lasts about 35 h [82]. Since a full simulation of the *in silico* tumors lasts about 800 cycles, the typical simulations reported here correspond to approximately 28 000 hours, about 38.9 months of real time.

The results shown in Fig. 9 indicate that for times longer than 12 months ($T > 300$ generations) the FD of the spatial distribution of the genes TP53, ATM, and BRCA1 become asymptotically closer to each other. In fact, the trend shown in the figure suggests that the FD behavior of the genes TP53, ATM, BRCA1, ERBB2, and MDM2 will asymptotically collapse onto one single curve for times $T > 800$ cycles, or longer 38.9 months of real time. Hence the time evolution of the FDs of the whole tumor and the spatial distribution of gene TP53 show similar quantitative behaviors. This result strongly indicates that the fractal structure of the whole tumor is fully determined by the spatial distribution of the tumor suppressor gene TP53 that plays a crucial role in cancer progression. On the other hand, the inset of Fig. 9 shows the time evolution of the whole tumor FD for five representative values of β. For $\beta \leqslant 1$, the FD follows approximately one single curve; however, for $\beta > 1$, the FD becomes systematically larger. This is not surprising since the larger the values of β, the

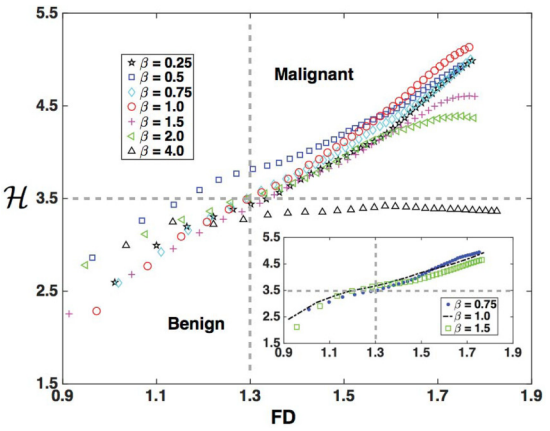

FIG. 10. Diversity $\mathcal{H}$ versus FD for the six genes set and several values of β. The results correspond to a tumor with parameters: $\alpha = 4 \times 10^{-3}$; $\lambda_N = 100$. The curves intersect at about (1.3,3.5). Points in the upper-right quadrant correspond to a malignant tumor, while those in the lower-left quadrant correspond to a benign tumor. The results shown in the inset correspond to the sixteen gene set and are consistent with those obtained with the set of six genes.

number of mutations decrease and, on the contrary, the smaller the values of β, the number of mutations increases. These results suggest that as proliferation increases (small values of β), the FD time evolution approaches a single universal curve.

In silico tumors have been generated for the model parameters $\alpha = 4 \times 10^{-3}, \lambda_N = 100$, different values of the parameter β, and the six-gene mutation dynamics. For each of them, the diversity, $\mathcal{H}$, and the FD have been calculated. With the results of these simulations, a "2D malignancy diagram," $\mathcal{H}$ versus FD, has been calculated. The results are shown in Fig. 10. There one sees that $\mathcal{H}$ increases monotonically as a function of FD for all values of β. Note that, with the exception of the curve corresponding to $\beta = 0.5$, all the other curves intersect at the crossing point of the dashed lines with coordinates $FD = 1.3$, $\mathcal{H} = 3.5$. This behavior suggests that there is a transition point in the tumor structure as a result of the spatial distribution of mutations. These results are consistent with recent studies that found that benign breast cancer tumors show a FD <1.3, while malignant tumors show a FD >1.3 [81]. Considering that $\mathcal{H} = 3.5$ is a transition point for diversity one can divide the plane $\mathcal{H}$ versus FD into four quadrants with the axis crossing point located at FD $= 1.3$, $\mathcal{H} = 3.5$ as indicated in Fig. 10. The (FD, $\mathcal{H}$) points located in the upper right quadrant FD > 1.3 and $\mathcal{H} > 3.5$ correspond to a malignant tumor. However, points located in the lower left quadrant (FD < 1.3 and $\mathcal{H} < 3.5$) correspond to a benign tumor. The inset of Fig. 10 shows the malignancy diagram obtained with a sixteen-gene mutation dynamics for three values of β. The diagram looks similar to that obtained with a six-gene dynamics, as shown in Fig. 10. All these results indicate that the predictions of our model are robust regardless of the use of six- or sixteen-gene mutation dynamics. In addition, a benign tumor is associated to a low fitness that is related to a high nutrient consumption while a malignant tumor is associated to a low fitness and a low nutrients consumption rate. The transition to the malignant state occurs as a consequence of a high nutrients consumption leaving spaces with low or no concentrations so that cells that are located there suffer starvation and eventually improve their fitness. This process leads to the appearance of bad mutations that leads to tumor growth. This diagram is consistent with previous observations [24,25].

V. DISCUSSION AND CONCLUSIONS

The present 2D model and its predictions can be contrasted with a 3D model that was proposed recently [39]. First, our model describes the growth of an avascular tumor with a cellular automata whose probabilities of division and death as well as the genetic dynamics depend on nutrients concentration gradients. By contrast, the 3D model combines spatial growth and accumulation of multiple mutations, while tumor growth is modeled by using an Eden lattice. Each cell in the tumor is described by its position and a list of genetic alterations that occur since the initial neoplastic cell. They focus on the interplay of genetics, spatial expansion, and short-range dispersal of cells, without including metabolism, tissue mechanics, spatial heterogeneity of tissues, different types of cells nor angiogenesis. Secondly, the present model uses a geometric probability distribution to enhance locally random mutations that are generated with a Poisson probability distribution. Notwithstanding that genetic alterations follow the same dynamics, the mutation accumulation per gene is described by a negative binomial distribution. In the division process of the 3D model each of the daughter cells receives new genetic alterations drawn from the Poisson probability distribution with parameter γ_x. Here the index refers to three possibilities: all mutations, driver mutations, or resistant mutations. These mutation probabilities are not known but they were varied in accordance to clinical data of different cancer types. In spite of the differences in spatial dimension and genetic dynamics, both models yield the hallmarks of cancer, spatial heterogeneity, and diversity. As a matter of fact, both models are able to describe some clinical aspects of tumor growth. More importantly, these models incorporate both spatial growth and genetic dynamics that previous models had considered separately.

In conclusion, we have presented and analyzed a quantitative growth model of an avascular tumor that considers the basic biological principles of cell proliferation, motility, death, transport of nutrients, and gene mutation dynamics. We postulate that the gene mutation rate depends on both randomness and microenvironmental factors, such as essential and nonessential nutrient concentrations for cell proliferation. It was found that higher concentrations of nutrients is an advantage that favors cancer cell proliferation as well as a high accumulation of mutations, which in turn leads to genetic diversity and tumor heterogeneity. Gene mutation dynamics considered two sets of genes, one with six genes from a regulatory network analysis and the other with sixteen genes, from an analysis of genomic data, which are believed to play a crucial role in cancer progression. The coupling of mutation dynamics to microenvironmental factors was done by introducing the parameter, β, that modulates proliferation rate together with a probability distribution that regulates mutation dynamics. The mutations in turn define the diversity and heterogeneity of the tumor. For $\beta < 1$, the mutation rate

is high and leads to high tumor gene diversity, whereas for $\beta > 1$, the mutation rate is low and the tumor diversity becomes normal.

For a given tumor one can calculate the diversity, $\mathcal{H}$, and the FD for different values of the parameter β. Thus a "malignancy diagram" based on $\mathcal{H}$ versus FD was calculated in Fig. 10. With the exception of the curve corresponding to $\beta = 0.5$, all the other curves intersect at the crossing point $FD = 1.3$, $\mathcal{H} = 3.5$, suggesting that this point indicates a critical change in the tumor behavior. More importantly, the results presented here suggest that the predictions of our model are robust whether we use six- or sixteen-gene sets for the mutation dynamics. In addition, our findings suggest that tumor fractal structure and diversity are fully determined by the heterogeneous spatial distribution of mutations of gene TP53, which is thought to play a crucial role in cancer progression. We would like to point out that the predictions of our model can be quantitatively related to clinical and experimental observations.

It is important to mention that the hallmarks of cancer, spatial heterogeneity and diversity, are reproduced in [39]. In spite of the fact that the present model is 2D it also yields the spatial heterogeneity and genetic diversity. These findings appear to suggest that there might be a universal mechanism independent of the tumor spatial dimension and genetic dynamics in cancer development.

ACKNOWLEDGMENTS

J.R.R.A. would like to acknowledge financial support from CONACyT-FORDECyT under Grant No. 265667. G.R.S. would like to acknowledge financial support from DGAPA-UNAM Grant No. IN108916. J.X.V.H. would like to acknowledge financial support from DGAPA-UNAM Grant No. IN110917. M.H.R. would like to acknowledge financial support from CONACYT-Ciencia Básica under Grant No. 54206. J.R.R.A. and G.R.S. would like to acknowledge enlightening discussions with Professor Enrique Hernández-Lemus. We would like to thank Carlos González-Castro for his technical support in the development of the stochastic simulations code. We would also like to thank Andres García-García for contributing to the genomic data search and sorting as well as a preliminary analysis. Finally, we would like to thank the LAVIS (Laboratorio Nacional de Visualizacin Cientfica Avanzada) from UNAM Juriquilla, Mexico, for the provided computer infrastructure, as well as Alberto Luis Aguilar for the technical support during the development of this project.

J.R.R.A., G.R.S., and J.X.H.V. developed the model. J.R.R.A. and G.R.S. carried out the simulations, the analysis of the results, and wrote the paper. L.O. developed the preliminary version of the code for the tumor growth and proofread the final version of the manuscript. M.H.R. provided the genomic data.

[1] P. M. Altrock, L. L. Liu, and F. Michor, The mathematics of cancer: integrating quantitative models, Nat. Rev. Cancer **15**, 730 (2015).

[2] *Driver and Passenger Mutations in Cancer*, edited by A. K. Abbas, S. J. Galli, and P. M. Howley, Annu. Rev. Path.-Mech. Disease **10** (2015).

[3] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr., and K. W. Kinzler, Cancer genome landscapes, Science **6127**, 1546 (2013).

[4] L. A. Garraway and E. S. Lander, Lessons from the cancer genome, Cell **1**, 17 (2013).

[5] M. R. Stratton, P. J. Campbell, and P. A. Futreal, The cancer genome, Nature (London) **458**, 719 (2009).

[6] A. G. Knudson, Two genetic hits (more or less) to cancer, Nat. Rev. Cancer **1**, 157 (2001).

[7] C. Carmona-Fontaine, M. Deforet, L. Akkari, C. B. Thompson, J. A. Joyce, and J. B. Xavier, Metabolic origins of spatial organization in the tumor microenvironment, Proc. Natl. Acad. Sci. USA **114**, 2934 (2017).

[8] A. Balmain, Cancer genetics: from Boveri and Mendel to microarrays, Nat. Rev. Cancer **1**, 77 (2001).

[9] K. L. Manchester, Theodor Boveri and the origin of malignant tumours, Trends Cell Biol. **5**, 384 (1995).

[10] C. M. Croce, Oncogenes and cancer, N. Engl. J. Med. **358**, 502 (2008).

[11] J. B. Konopka, S. M. Watanabe, J. W. Singer, S. J. Colins, and O. N. White, Cell lines and clinical isolates derived from Ph1-positive chronic myelongenous leukemia patients express c-abl proteins with common structural alteration, Proc. Natl. Acad. Sci. USA **82**, 1810 (1985).

[12] Y. Tsujimoto, J. Gorham, J. Cossman, E. Jaffe, and C. M. Croce, The t(14;18) chromosome translocations involved in b-cells neoplasms results from mistakes, Science **229**, 1390 (1985).

[13] L. R. Finger, R. C. Harvey, R. C. A. Moore, and C. M. Croce, A common mechanism of chromosomal translocation in t and b-cells neoplasia, Science **234**, 982 (1986).

[14] J. M. Bishop, Molecular themes in oncogenesis, Cell **44**, 235 (1991).

[15] I. Dagogo-Jack and A. T. Shaw, Tumour heterogeneity and resistance to cancer therapies, Nat. Rev. Clin. Oncol. **15**, 81 (2017).

[16] A. Heindl, S. Nawaz, and Y. Yuan, Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology, Lab. Invest. J. Tech. Methods Pathol. **4**, 377 (2015).

[17] P. C. Nowell, The clonal evolution of tumor cell populations, Science **4260**, 23 (1976).

[18] M. Greaves and C. C. Maley, Clonal evolution in cancer, Nature (London) **7381**, 306 (2012).

[19] N. McGranahan, F. Favero, E. C. de Bruin, N. J. Birkbak, Z. Szallasi, and C. Swanton, Clonal status of actionable driver events and the timing of mutational processes in cancer evolution, Sci. Transl. Med. **7**, 283 (2015).

[20] Y. Yuan, Y. Jiang, C. Sun, and Q. Chen, Role of the tumor microenvironment in tumor progression and the clinical applications (review), Oncol. Rep. **35**, 2499 (2016).

[21] R. J. DeBerardinis and N. S. Chandel, Fundamentals of cancer metabolism, Sci. Adv. **2**, e1600200 (2016).

[22] C. A. Lyssiotis and A. C. Kimmelman, Metabolic interactions in the tumor microenvironment, Trends Cell Biol. **27**, 863 (2017).

[23] J. Espinal-Enríquez, C. Fresno, G. Anda-Jáuregui, and E. Hernández-Lemus, RNA-Seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer, Sci. Rep. **7**, 1760 (2017).

[24] J. Yuan and P. M. Glazer, Mutagenesis induced by the tumor microenvironment, Mutat. Res., Fundam. Mol. Mech. Mutagen. **400**, 439 (1998).

[25] A. S. Lee, Coordinated regulation of a set of genes by glucose and calcium ionophores in mammalian cells, Trends Biochem. Sci. **12**, 20 (1987).
[26] J. Haigh, The accumulation of deleterious genes in a population—Muller's ratchet, Theor. Pop. Biol. **14**, 251 (1978).
[27] N. N. Pavlova and C. B. Thompson, The emerging hallmarks of cancer metabolism, Cell Metab. **23**, 27 (2016).
[28] D. Anastasiou, Tumour microenvironment factors shaping the cancer metabolism landscape, Br. J. Cancer **116**, 277 (2016).
[29] D. Drasdo and S. Höhme, A single-cell-based model of tumor growth in vitro: monolayers and spheroids, Phys. Biol. **2**, 133 (2005).
[30] T. Roose, S. Chapman, and P. Maini, Mathematical models of avascular tumor growth, SIAM Rev. **49**, 179 (2007).
[31] Y. Jiang, J. Pjesivac-Grbovic, C. Cantrell, and J. P. Freyer, A multiscale model for avascular tumor growth, Biophys. J. **89**, 3884 (2005).
[32] T. Alarcón, H. Byrne, and P. Maini, A multiple scale model for tumor growth, Multiscale Model. Simul. **3**, 440 (2005).
[33] J. Jeon, V. Quaranta, and P. T. Cummings, An off-lattice hybrid discrete-continuum model of tumor growth and invasion, Biophys. J. **98**, 37 (2010).
[34] J. A. Sherratt and M. A. Chaplain, A new mathematical model for avascular tumour growth, J. Math. Biol. **43**, 291 (2001).
[35] R. A. Gatenby and E. T. Gawlinski, A reaction-diffusion model of cancer invasion, Cancer Res. **56**, 5745 (1996).
[36] G. Albano and V. Giorno, A stochastic model in tumor growth, J. Theor. Biol. **242**, 329 (2006).
[37] V. Cristini, X. Li, J. S. Lowengrub, and S. M. Wise, Nonlinear simulations of solid tumor growth using a mixture model: invasion and branching, J. Math. Biol. **58**, 723 (2008).
[38] D. Bresch, T. Colin, E. Grenier, B. Ribba, and O. Saut, Computational modeling of solid tumor growth: The avascular stage, SIAM J. Sci. Comput. **32**, 2321 (2010).
[39] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak, A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity, Nature (London) **525**, 261 (2015).
[40] S. C. Ferreira, M. L. Martins, and M. J. Vilela, Reaction-diffusion model for the growth of avascular tumor, Phys. Rev. E **65**, 021907 (2002).
[41] J. Paulsson, O. G. Berg, and M. Ehrenberg, Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation, PNAS **97**, 7148 (2000).
[42] N. Beerenwinkel, T. Antal, D. Dingli, A. Traulsen, K. W. Kinzler, V. E Velculescu, B. Vogelstein, and M. A. Nowak, Genetic progression and the waiting time to cancer, PLoS Comput. Biol. **3**, 1 (2007).
[43] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[44] M. A. Medina and I. Nunez de Castro, Glutaminolysis and glycolysis interactions in proliferant cells, Int. J. Biochem. **22**, 681 (1990).
[45] L. Wang, J.-J. Li, L.-Y. Guo, P. Li, Z. Zhao, H. Zhou, and L.-J. Di, Molecular link between glucose and glutamine consumption in cancer cells mediated by CtBP and SIRT4, Oncogenesis **7**, 26 (2018).
[46] A. B. Herman, V. M. Savage, and G. B. West, A quantitative theory of solid tumor growth, metabolic rate and vascularization, PLoS ONE **6**, e22973 (2011).
[47] K. Groebe and W. Mueller-Klieser, On the relation between size of necrosis and diameter of tumor spheroids, Int. J. Radiat. Oncol., Biol., Phys. **34**, 395 (1996).
[48] C. Yu and J. Wang, A physical mechanism and global quantification of breast cancer, PLoS ONE **11**, e0157422 (2016).
[49] J. R. Pon and M. A. Marra, Driver and passenger mutations in cancer, Annu. Rev. Pathol.: Mech. Dis. **10**, 25 (2015).
[50] I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak, Accumulation of driver and passenger mutations during tumor progression, Proc. Natl. Acad. Sci. USA **107**, 18545 (2010).
[51] I. Martincorena, K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, H. Davies, M. R. Stratton, and P. J. Campbell, Universal patterns of selection in cancer and somatic tissues, Cell **171**, 1029 (2017).
[52] M. D. Robinson and G. K. Smyth, Moderated statistical tests for assessing differences in tag abundance, Bioinformatics **23**, 2881 (2007).
[53] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics **26**, 139 (2010).
[54] N. Weinhold, A. Jacobsen, N. Schultz, C. Sander, and W. Lee, Genome-wide analysis of noncoding regulatory mutations in cancer, Nat. Genet. **46**, 1160 (2014).
[55] E. T. Wang, R. Sandberg, L. Shunjun, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge, Alternative isoform regulation in human tissue transcriptomes, Nature (London) **456**, 470 (2008).
[56] Z. Fang, J. Martin, and Z. Wang, Statistical methods for identifying differentially expressed genes in RNA-Seq experiments, Cell Biosci. **2**, 26 (2012).
[57] A. Anjum, S. Jaggi, E. Varghese, S. Lall, A. Bhowmik, and A. Rai, Identification of differentially expressed genes in RNA-Seq data of arabidopsis thaliana: A compound distribution approach, J. Comput. Biol. **23**, 239 (2016).
[58] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications* (McGraw-Hill, New York, 1997).
[59] R. Gatenby and R. J. Gillies, A microenvironmental model of cancinogenesis, Nat. Rev. Cancer **8**, 56 (2008).
[60] R. Durrett, Population genetics of neutral mutations in exponentially growing cancer cell populations, Ann. Appl. Probab. **23**, 230 (2013).
[61] C. D. McFarland, K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny, Impact of deleterious passenger mutations on cancer progression, Proc. Natl. Acad. Sci. USA **110**, 2910 (2013).
[62] C. Sala, S. Vitali, E. Giampieri, Ì. Faria do Valle, D. Remondini, P. Garagnani, M. Bersanelli, E. Mosca, L. Milanesi, and G. Castellani, Stochastic neutral modeling of the gut microbiota's relative species abundance from next generation sequencing data, BMC Bioinform. **17**, 16 (2016).
[63] D. T. Gillespie, Stochastic simulation of chemical kinetics, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[64] Y. Cao, D. T. Gillespie, and L. R. Petzold, Avoiding negative populations in explicit poisson tau-leaping, J. Chem. Phys. **123**, 054104 (2005).

[65] T. T. Marquez-Lagoa and K. Burrageb, Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics, J. Chem. Phys. **127**, 104101 (2007).

[66] M. F. Pettigrew and H. Resat, Multinomial tau-leaping method for stochastic kinetic simulations, J. Chem. Phys. **126**, 084101 (2007).

[67] J. Pahle, Biochemical simulations: stochastic, approximate stochastic and hybrid approaches, Brief. Bioinf. **10**, 53 (2009).

[68] H. Holten-Rossing, M.-L. Møller Talman, A. M. B. Jylling, A.-V. Lænkholm, M. Kristensson, and B. Vainer, Application of automated image analysis reduces the workload of manual screening of sentinel lymph node biopsies in breast cancer, Histopathology **71**, 866 (2017).

[69] S. Y. Nam, E. Y. Ko, B.-K. Han, J. H. Shin, E. S. Ko, and S. Y. Hahn, Breast imaging reporting and data system category 3 lesions detected on whole-breast screening ultrasound, J. Breast Cancer **19**, 301 (2016).

[70] H. Raza Ali, A. Dariush, E. Provenzano, H. Bardwell, J. E. Abraham, M. Iddawela, A.-L. Vallier, L. Hiller, J. A. Dunn, S. J. Bowden, T. Hickish, K. McAdam, S. Houston, M. J. Irwin, P. D. P. Pharoah, J. D. Brenton, N. A. Walton, H. M. Earl, and C. Caldas, Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer, Breast Cancer Res.: BCR **18**, 21 (2016).

[71] J. Vasiljevic, J. Pribic, K. Kanjer, W. Jonakowski, J. Sopta, D. Nikolic-Vukosavljevic, and M. Radulovic, Multifractal analysis of tumour microscopic images in the prediction of breast cancer chemotherapy response, Biomed. Microdevices **17**, 93 (2015).

[72] A. David and S. Vassilvitskii, *K-means++: The advantages of careful seeding, SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007 (unpublished), p. 1027.

[73] A. E. Magurran, *Measuring Biological Diversity* (Wiley-Blackwell, Hoboken, NJ, 2004).

[74] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory **28**, 129 (1982).

[75] S. J. Potts, J. S. Krueger, N. D. Landis, D. A. Eberhard, G. David Young, S. C. Schmechel, and H. Lange, Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue, Lab. Investigation **92**, 1342 (2012).

[76] S. Y. Park, M. Gönen, H. J. Kim, F. Michor, and K. Polyak, Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype, J. Clin. Invest. **120**, 636 (2010).

[77] V. Almendro, H. J. Kim, Y.-K. Cheng, M. Gönen, S. Itzkovitz, P. Argani, A. van Oudenaarden, S. Sukumar, F. Michor, and K. Polyak, Genetic and phenotypic diversity in breast tumor metastases, Cancer Res. **74**, 1338 (2014).

[78] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang, Pathology imaging informatics for quantitative analysis of whole-slide images, J. Am. Med. Inf. Assoc.: JAMIA **20**, 1099 (2013).

[79] M. Veta, J. P. W. Pluim, P. J. Van Diest, and M. A. Viergever, Breast cancer histopathology image analysis: A review, IEEE Trans. Biomed. Eng. **61**, 1400 (2014).

[80] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, A minimum spanning forest based method for noninvasive cancer detection with hyperspectral imaging, IEEE Trans. Bio-med. Eng. **63**, 653 (2016).

[81] R. Dobrescu, I. Ichim, and D. Crişan, Diagnosis of breast cancer from mammograms by using fractal measures, Int. J. Biomed. Imag. **1**, 32 (2013).

[82] S. Larsson, T. Rydén, U. Holst, S. Oredsson, and M. Johansson, Estimating the total rate of dna replication using branching processes, Bull. Math. Biol. **70**, 2177 (2008).