

Quantitative Interpretation of a Genetic Model of Carcinogenesis Using Computer Simulations

Donghai Dai^{1*}, Brandon Beck¹, Xiaofang Wang¹, Cory Howk¹, Yi Li²

1 Department of Obstetrics & Gynecology, University of Iowa, Iowa City, Iowa, United States of America, **2** Department of Mathematics, University of Iowa, Iowa City, Iowa, United States of America

Abstract

The genetic model of tumorigenesis by Vogelstein et al. (V theory) and the molecular definition of cancer hallmarks by Hanahan and Weinberg (W theory) represent two of the most comprehensive and systemic understandings of cancer. Here, we develop a mathematical model that quantitatively interprets these seminal cancer theories, starting from a set of equations describing the short life cycle of an individual cell in uterine epithelium during tissue regeneration. The process of malignant transformation of an individual cell is followed and the tissue (or tumor) is described as a composite of individual cells in order to quantitatively account for intra-tumor heterogeneity. Our model describes normal tissue regeneration, malignant transformation, cancer incidence including dormant/transient tumors, and tumor evolution. Further, a novel mechanism for the initiation of metastasis resulting from substantial cell death is proposed. Finally, model simulations suggest two different mechanisms of metastatic inefficiency for aggressive and less aggressive cancer cells. Our work suggests that cellular de-differentiation is one major oncogenic pathway, a hypothesis based on a numerical description of a cell's differentiation status that can effectively and mathematically interpret some major concepts in V/W theories such as progressive transformation of normal cells, tumor evolution, and cancer hallmarks. Our model is a mathematical interpretation of cancer phenotypes that complements the well developed V/W theories based upon description of causal biological and molecular events. It is possible that further developments incorporating patient- and tissue-specific variables may build an even more comprehensive model to explain clinical observations and provide some novel insights for understanding cancer.

Citation: Dai D, Beck B, Wang X, Howk C, Li Y (2011) Quantitative Interpretation of a Genetic Model of Carcinogenesis Using Computer Simulations. PLoS ONE 6(3): e16859. doi:10.1371/journal.pone.0016859

Editor: Ying Xu, University of Georgia, United States of America

Received: August 30, 2010; **Accepted:** January 14, 2011; **Published:** March 8, 2011

Copyright: © 2011 Dai et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by ACS Research Scholar Award RSG-06-105-01-CCE (DD) and UI College of Medicine Basic Reproductive Science Research Fund. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: donghai-dai@uiowa.edu

Introduction

Efforts to understand cancer have continued to intensify since the start of a presidential campaign to conquer cancer in 1971 [1]. Mortality rates from cancer remains stubbornly high with more than half a million deaths in the US alone in 2009 [2]. Cancer is considered to be predominantly a genetic disease [3]. It is believed that multiple sequential mutations induce malignant transformation of a normal cell into the cancer founder cell, which then multiplies and evolves to become a clinically detectable tumor [4,5]. This genetic model of carcinogenesis (referred to here as V theory) is augmented by the elegant description of major cancer features by Weinberg et al which is recognized as a seminal and most comprehensive molecular definition of cancer [3,6]. We specify these two seminal models collectively as the V/W theory of cancer and have sought to develop a mathematical model capable of quantitatively interpreting the V/W theory.

Epithelial tissues are the most common locus of oncogenesis. However, individual epithelial cells are in a constant developmental process of tissue regeneration, namely from stem cell to proliferating/differentiating cell and, finally, to senescent cells [7]. The short lifetime and continued proliferation of epithelial cells in a tissue with a population of 10^{12} cells pose numerous challenges to determining the natural course of oncogenesis. We have,

therefore, attempted to describe the life cycle of an epithelial cell clone of endometrial origin as a normal physiological process to serve as a basic reference for oncogenesis, which is made possible by the addition of many genetic and environmental factors.

Clone lifetimes during normal epithelial cell regeneration and some major assumptions for the study of carcinogenesis

Endometrial cancer arises in the uterine epithelium, which even in adults are undergoing constant turnover. The tissue stem cells provide a stable cell source for tissue regeneration [7]. A stem cell produces a progenitor cell committed to proliferation, resulting in a clone with hundreds of descendant cells through many generations of cell division. If we assume that the tissue stem cell pool will provide as many progenitor cells as it needs at any time to ensure tissue homeostasis (a stable total cell number), the life span of a clone, from zero cell number (before the birth of a progenitor cell) to one progenitor, to hundreds of descendant cells, to senescence and eventually death, is a cycle from zero cells at the beginning to zero cells at the end over a short time period, days or months. A mathematical description of normal tissue regeneration may identify immortalization (defying programmed senescence and cell death) as an early deviation from the physiological process

with a potential for oncogenesis as a result from a combined effect of genetic alterations and environmental stimulations [8]. It thus creates mathematically a continuous and wide spectrum of physiological and pathological cellular events with cancer at the other end. The progenitor cell immediately born from a stem cell and its descendant cells are defined as a **clone** and the entire process (from 0 to 0) is defined as the **clone lifetime**. Thus, any non-stem cell can be tagged and quantitatively analyzed according to chronological time during a clone lifetime. Stem cells are excluded for calculation since their number is maintained through symmetric and asymmetric divisions in the stem cell compartment. For uterine epithelium, there are millions of clones which are actively cycling at any moment but at different stages of their clone lifetime. This ensures a stable total cell number in a tissue for tissue homeostasis.

The progression of a clone lifetime from the progenitor cell is not only shown in the increase of cell number, but also in accompanying differentiation. We derive a mathematical expression to describe the kinetics of a single cell during a clone lifetime. Consequently, a tissue or any population of cells could be treated as a composite of individual cells with cell-specific variables.

Cellular proliferation can be described in terms of density-dependent growth $\frac{dN}{dt} = f(t, N)N$ for some rate function $f(t, N)$. Tumor growth is often studied in the context of Gompertzian growth, where $f(t, N) = -b \log(\frac{N}{K})$, for constants b and K [9,10]. However, the Gompertz model is empirical and is based on data-fitting with respect to tumor volume and weight for its description of tumor growth [9,10], although mathematical explanations for the appearance of this growth term have been proposed. In our model we instead consider tumor size as a result of the proliferation of its constituent cells. We derive a set of equations to describe the lifespan of an individual cell within the mass. This process allows us to describe intraclonal heterogeneity in response to mutations and environmental stimulation.

Cellular proliferation is described through a cell's proliferation potential (α , with a unit of doublings per unit time) [11], and occurs by the following:

$$n(t) = \int_{t_n}^t \alpha(s) ds, \tag{1}$$

where $\alpha(t)$ is a cell's proliferation potential and t_n is the time the cell was born. The function $n(t)$ can be interpreted as a measure of a cell's status toward either dividing or dying. If $n(t_*) = 1$ for some time $t = t_*$ the cell divides, while if $n(t_*) = -1$ the cell dies. The equation for $n(t)$ is a consequence of the more common growth terms

$$\frac{dN}{dt} = \ln(2)\alpha(t)N(t), \text{ or equivalently, } N(t) = 2^{\int_{t_n}^t \alpha(s) ds}, \tag{2}$$

which can be confusing in this context since we are discussing the proliferation of a single cell and not an entire population. The instantaneous α value for any cell in a tumor at any time of its history is measurable theoretically and can be approximated in practice with a mean value which is measured by the change of cell number over a short time period [12]. Hayflick et al described an individual cell's capability to multiply as doubling potential [13]. This concept was further expressed as proliferation potential [14,15]. It was believed to be critically important to describe intraclonal heterogeneity through description of the difference of doubling potential of individual cells [13]. Consequently, intraclonal heterogeneity can be quantitatively described in a

population (a tissue or a tumor) through the summation of individual cells even though all equations here are derived to describe a single cell.

A clone lifetime, starting from a progenitor cell immediately born from a stem cell can be mathematically expressed to capture a dynamic interplay of proliferation and differentiation:

$$\alpha_p(t) = (1/L) \cdot (L - g(t)) \cdot g(t) \tag{3}$$

$$k_p(t) = K_{max} \cdot [1 - e^{-K' \cdot g(t)}] \tag{4}$$

$$g(t) = \text{ceiling} \left(\int_0^t |\alpha(s)| ds \right) \tag{5}$$

where $g(t)$ is the generation (g) at time t (a quantitative measurement of lineage progression, related to cell division), $\alpha_p(t)$ is the generation-dependent programmed (inherent) α , and $k_p(t)$ is the corresponding programmed differentiation coefficient (k). The terms L , L' , K_{max} and K' are tissue type-specific constants. Generally, L is designated as the maximal programmed generation number before a cell enters a senescent phase according to a cell's inherent developmental program. L' is the tissue-specific conversion coefficient, which converts mathematically the value of generation to that of proliferation potential, dictating the change of programmed proliferation potential value according to a cell's biological clock, a reading of progression of generation. K_{max} is the coefficient representing the maximal value of differentiation in a tissue for the most mature cell, where greater differentiation implies greater homeostasis according to Assumption 1 below. K' is the respective tissue-specific differentiation conversion coefficient.

Proliferation and differentiation are integrated and continual cell replacement processes in adult tissue homeostasis [7,16]. In our model, α_p during the developmental process can be described as a gradual increase in its value, reaching the peak at the middle of the developmental stage, and then undergoing a gradual decrease until senescence. Meanwhile, k_p increases gradually and reaches a plateau (maximal level) when cells become well-differentiated. The idea is that cells at the middle of the clone lifetime are the most productive and specialization continues throughout the lifetime. When $g > L$, α_p becomes negative, indicating that the cell should enter a process of senescence by its inherent program. This cellular dynamic mimics the gradual proliferation and differentiation process in uterine epithelium during the menstrual cycle.

Here, we must introduce the following assumptions in order to describe the transformation process of oncogenesis.

Assumption 1. We hypothesize that one of the most important properties of a normal cell is the maintenance of a programmed proliferation potential α_p in a developmental stage-specific manner: homeostasis in proliferation or cell number. We define a cell's ability to maintain α_p as the resistance potential (r) with the following.

$$r(t) = k_p(t) \cdot (\alpha_p(t) - \alpha(t)), \tag{6}$$

where $r(t)$ is the resistance potential and $\alpha(t)$ is the proliferation potential at time t . The resistance potential works to bring a cell's α to its programmed level α_p , and thus is defined as the force to maintain homeostasis in proliferation. Resistance potential could be executed by cell cycle regulators such as cyclin D1, cdk1, Rb,

p16, p27 and p53, one of the most important players capable of arresting cell cycle or triggering apoptosis.

Assumption 2. A human cell has in excess of 3 billion base pairs in its genome. Genetic alterations (or mutations), including single and multiple base pair changes, chromosome translocation, aneuploidy and epigenetic alterations, are broadly defined here as any altered hereditary factor, with an enormous number of possibilities. A genetic alteration's effect m_i on differentiation status (numerically expressed as differentiation coefficient k) is given as the following:

$$k(t) = k_p(t) + \sum_{i=1}^n m_i \quad (7)$$

where n is the number of genetic alterations accrued by the cell during the proliferative process, and m_i is the numerical value of a genetic alteration, quantifying its effect on the differentiation coefficient. The values m_i follow a Gaussian distribution with mean μ and standard deviation σ as $m_i \sim N(\mu, \sigma)$. It has been reported that differentiation status of a tumor can be quantitatively defined and tumor subtypes can be classified accordingly [17]. Thus, quantitative change of tumor differentiation should arise from accumulation of mutations according to the V theory of tumorigenesis [4,5].

Assumption 3. A cell lives in a microenvironment filled with various growth and anti-growth signals such as hormones, growth factors, and cytokines. Spatial constraints could have a remarkable effect on cell proliferation as well, especially when a cell is proliferating rapidly (in the case of cancer [18,19]). The specific effect on a cell by a particular signal is not fixed and not precisely targeted. We assume that the collective effect of all signals on a particular cell follows a Gaussian distribution: $\beta \sim N(\mu, \sigma)$. The effect of β on cell proliferation is expressed by a dynamical change in α . This effect, along with the resistance potential, combines to alter α according to:

$$\frac{d\alpha}{dt} = r(t) + \beta(t). \quad (8)$$

Instead of combining the assumptions above into a single equation, we present the following four parts collectively as a summary equation for the growth dynamics of an individual cell in order to discuss them more intuitively:

$$n(t) = \int_m^t \alpha(s) ds, \quad \frac{d\alpha}{dt} = r(t) + \beta(t), \quad r(t) = k(t) \cdot (\alpha_p(t) - \alpha(t)), \quad (9)$$

$$k(t) = k_p(t) + \sum_{i=1}^n m_i.$$

The clone size is then determined by the summation of all cells within the clone growing according to these dynamics.

Results

In this section, we will explore tissue homeostasis, malignant transformation, tumor evolution, an alternate mechanism for the initiation of metastasis and the survival and establishment of metastatic lesions. The focus will be on endometrial cancer. The mathematical description of normal tissue regeneration is essential to serve as a reference for the process of malignant transformation. Epithelial cell turnover is a dynamic process involving billions of

cells and is presented here as a huge landscape with genetic insults (lightning) and environmental stimulations (raining) having an uneven effect on individual cells which will consequently serve as the basis of heterogeneity of oncogenesis and tumor evolution (Fig. 1).

The clone lifetime of uterine epithelial cells

We first simulated the developmental process of normal epithelial cells in order to understand clone size, the lifespan of every descendant cell, proliferation potential at any moment of development, and more importantly, changes in differentiation coefficient. For the simulations of uterine epithelium proliferation, we set $L = 10$, $L' = 7$, $K_{max} = 3.78$, and $K' = 0.4$ in equations 3 and 4, yielding the following system to describe an individual cell in a clone lifetime:

$$\alpha_p(t) = (10 - g(t)) \cdot g(t) / 7 \quad (10)$$

$$k_p(t) = 3.78 \cdot [1 - e^{(-0.4g(t))}]. \quad (11)$$

For the purpose of a more simplified and specific discussion, we will use equations 10 and 11 to substitute for equations 3 and 4 in all of the simulations presented in this paper.

Since endometrial cancer primarily occurs in postmenopausal women, simulations of the normal developmental process in uterine endometrium were conducted either in the absence of estrogenic stimulation or in the presence of low amounts of estrogen exposure. This is represented numerically by either a neutral mean or a small positive mean value for the Gaussian distribution for β , respectively. Numerical simulations, with the specified values for L , L' , K_{max} , and K' , suggest that the possible maximal clone size from a single progenitor cell is about $2^9 = 512$ (Fig. 2a). However, when either a neutral or small positive environmental stimulation (β) is applied which acts randomly and independently on individual cells every day, cell division in a clone is no longer synchronized due to the differing $\hat{1} \pm$ values for each cell. Peak clone size and time for complete death of a clone vary according to β values. In a hypothetical menopausal endometrium without any hormonal effect (zero environment, Fig.2a), a clone will remain at a stable cell number for many years before they enter senescence. With a neutral environment, $\beta \sim N(0,1)$ or $\beta \sim N(0,2)$, the life cycle of a clone was remarkably shortened although clones took a similar amount of time to reach peak clone size (Fig. 2a). When a positive environment, $\beta \sim N(1,1)$ or $\beta \sim N(2,1)$, was applied, clone size increased dramatically and the time required to reach the peak clone size and complete senescence were substantially shortened (Fig. 2a).

The values of α and β which were set around $1/month$ and $1/month^2$, respectively, for a menopausal endometrium may seem too low since many cancer cells double in 1–2 days in culture. However, cancer cells are grown in strong stimulation in vitro with 10–20% fetal bovine serum. Our simulation thus far considers conditions without or with a minimal environmental stimulation. In order to get a sense for what a menstrual cycle will look like in the term of cell proliferation, a strong β , $\beta \sim N(60,6)$ or $\beta \sim N(90,9)$ or $\beta \sim N(120,12)$, was used to simulate a monophasic oral contraceptive where hormones were given for 21 days (Fig. 2b). The clone size could be 10–1,000 times larger depending upon the hormone levels, consistent with the observation of hyperplasia during proliferative phase of a menstrual cycle. Cell death was sudden when the hormones were withdrawn. However,

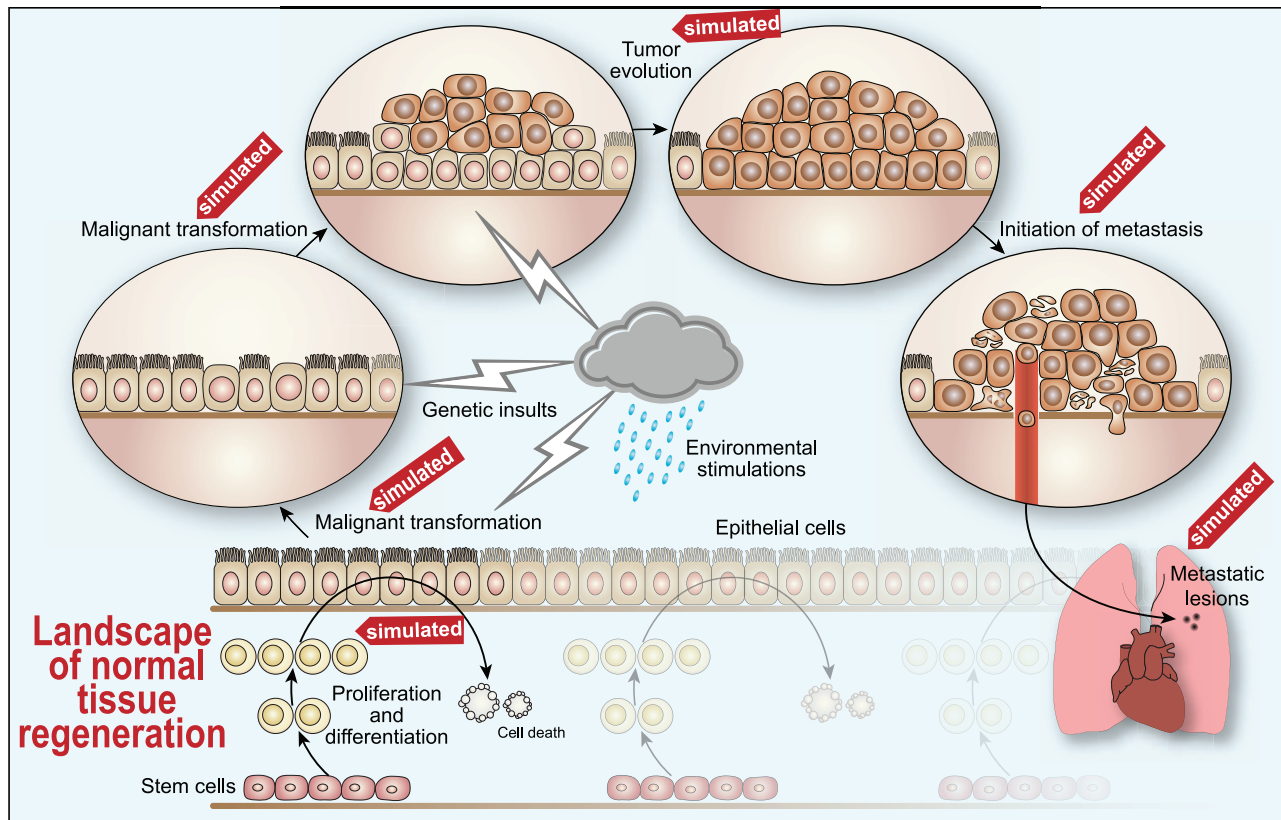


Figure 1. Schematic illustration of some of critical steps of human oncogenesis. These steps are simulated by our model to interpret prevailing genetic theories of cancer. Our model, consisting of a few simple equations, captures the landscape of normal uterine tissue regeneration (simulated for menopausal uterine epithelium and menstrual cycle in Fig. 2), malignant transformation (reduction of differentiation coefficient by genetic alterations in Table 1), occurrence of dormant/transient tumors (simulation of tumor incidence in Tables 2 and 3), tumor evolution (one cascade of tumor evolution resulting in the selection of more aggressive cancer subpopulations in Fig.3), a potential mechanism of the initiation of metastasis (simulated in Fig.4) and the development of clinically detectable metastatic lesions (simulated in Fig. 5). This model describes the evolution of individual cells incorporating quantitative effects from genetic alterations and environment factors, and emphasizes the role of intra-tumor heterogeneity in tumor evolution and metastasis.
doi:10.1371/journal.pone.0016859.g001

a clone lifetime was still longer than 28 days since we here did not incorporate any interaction between fast growth of cells and environment, which is simulated in a following section, and the pharmacological effect of sudden hormone withdrawal on blood vessels supporting the endometrial tissues, which is not considered in this manuscript. Once again, the entire endometrial tissue is composed of millions of clones at different stages of their clone lifetime. Every clone starts with zero number and finished with zero number over its entire clone lifetime, with tissue stem cells producing progenitor cells at any moment and resulting in millions of clones at active cycling. The stem cell compartment will not be included in our calculations in this manuscript.

Malignant transformation during dynamic epithelial cell development

Based on the above description of a clone lifetime, *malignant transformation* (the process of a normal cell becoming a cancer cell), is thus defined on the single cell level as a reduction of a cell's differentiation coefficient k (a process of de-differentiation). This is a definition at the single cell level as opposed to the definition of *tumor* or *cancer* in mass or tissue level. In the traditional definition, a *tumor* is defined as a clinically detected mass, a *cancer cell* is defined morphologically by pathological examination, and *malignancy* as its statistical correlation with patient outcome.

In order to simulate the potential effect of mutations on malignant transformation, one genetic alteration is assumed to occur at every cell division based on data available in the literature [5]. Every alteration occurs randomly and independently, and has an effect on the differentiation coefficient following a Gaussian distribution of $m \sim N(\mu, \sigma)$. In order to explore the quantitative link between the high incidence of transient and dormant tumors, and the mutation rate and environmental stimulations, we set $\mu = 0$ and $\sigma = 0.1$, consistent with the hypothesis that mutations could have either a positive or negative effect on the differentiation status, and that most mutations are passenger mutations with little effect.

Accumulation of genetic alterations should be discussed in a setting that a cell has a short lifetime physiologically in an epithelium ranging from days to months but the tissue has billions of cells at any moment and many times more in a human's lifetime. Clearly, a cell accumulates genetic alterations during the developmental process, with a well-differentiated cell (long living with $g > 10$) having the most alterations. Using the above values of μ and σ , and with the specified values for L , L' , K_{max} and K' , the probability for a single cell to have a $k=0$ (complete de-differentiation and thus malignant transformation) after 10 generations (11 mutations) is 4×10^{-30} (Table 1). Even assuming 10^{10} as the total number of cells in an epithelial tissue at any moment, and a one year lifetime of an epithelial clone (a rough

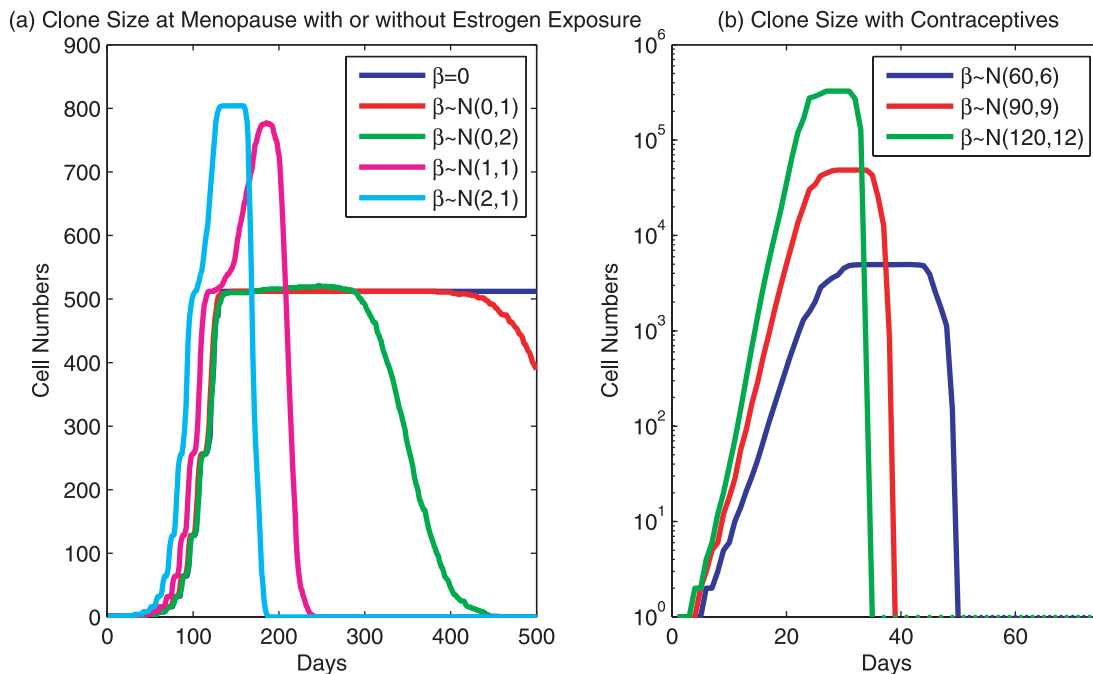


Figure 2. A clone life of the uterine epithelial cells. The life of a clone is perceived to start when a progenitor cell is borne from an asymmetrical division of a stem cell at $t=0$. The progenitor cell multiplies to produce the clone (all descendant cells), whose cells exhibit developmental stage specific features described by equations 10, 11 and 5. The clone size is calculated as the summation of all individual cells whose growth is governed by system 9. The size of the entire uterine epithelium at any time will be the summation of all clones with different sizes. A β value is assigned to an existing cell every day, randomly and independently. **a. Simulations of postmenopausal endometrium.** $\beta=0$ is used to simulate a hypothetical scenario exhibiting no environmental exposure. Neutral β 's, $\beta \sim N(0,1)$ and $\beta \sim N(0,2)$, are used as representations of the microenvironmental effect of postmenopause. A small positive β , $\beta \sim N(1,1)$ and $\beta \sim N(2,1)$, is introduced to reflect a chronic exposure to weak estrogenic stimulation. **b. Simulations of endometrium with monophasic contraceptives.** A monophasic schedule of oral contraceptives is used to model the menstrual cycle. Strong and positive β 's, $\beta \sim N(60,6)$, $\beta \sim N(90,9)$, and $\beta \sim N(120,12)$, are applied for 21 days in the simulations for the representation of oral estrogen and progestin. The interaction between fast cell growth and microenvironment, and the effect of sudden withdrawal of hormones on blood vessels are not included.
doi:10.1371/journal.pone.0016859.g002

estimate for postmenopausal uterine epithelium), a 100 year old woman has statistically no chance to harbor a fully transformed cell through genetic alterations alone. The probability to get a cell with $k=0$ is even smaller in a less-differentiated cell ($g < 10$) since its accumulated mutational load is small although its k value is low. Although tissue turnover could be as short as several days in some epithelia (e.g., colon crypts), a hundred fold increase in total cell number in an individual's lifetime could not overcome the rare probability of complete malignant transformation by mutation alone. Our model suggests that environmental factors and evolution could play critical roles in the selection and immortalization of partially transformed cells.

Tumor evolution: growth of a heterogeneous mass consisting of cells with various differentiation coefficients

In order to simulate the evolution of a heterogeneous tumor, we assumed that 4 masses have 1000 cells each at time $t=0$. Every cell had an $\alpha_p=0$ and a starting $\alpha(0) \sim N(0,1)$. A neutral and slightly varying environmental effect β was assumed to follow $\beta \sim N(0,1)$. We further assumed the distribution of k values in a mass followed $k \sim N(0,0)$, $(2,1)$, $(3,1)$ and $(4,1)$, respectively. The median mass with $k=0$ reached 10^{12} cells (the simulation limit when a simulation was terminated) by 11 months (Fig.3a). The median mass with $k \sim N(2,1)$ reached the simulation limit by 15 months (Fig.3a). The median size of masses with $k \sim N(3,1)$ was more than 10^8 cells by 24 months (Fig.3a). In the contrast, almost all cells in the mass with $k \sim N(4,1)$ would have a strong resistance

potential to maintain an $\alpha_p=0$ due to high k values (a possible least k value is about 1 according to the accumulated probability), and a neutral growth stimulation could not produce any significant change of tumor mass (Fig. 3a). While different values for L , L' , K_{max} , and K' change the times when the simulations reach a simulation limit, the overall dynamics remain the same.

For the mass with the original $k \sim N(2,1)$, the majority of cells with high k values did not grow much since their resistance potential quickly neutralizes the initial proliferation potentials ($\alpha(0)$) and any growth stimulatory effect from growth signals (β) to maintain $\alpha(t)$ around α_p level. It was cells with low k values that lost their capability to maintain $\alpha(t)$ at α_p level and multiplied remarkably, contributing increasingly to the population of the mass. The median k value of the mass declined quickly to reach $k=0$ by 6 months (Fig. 3b), indicating that the overwhelming portion of cells were those with $k=0$. Similarly, the median k value reached 0.0337 by 21 months for the mass with the original $k \sim N(3,1)$, which is approximately the lowest value among the original 1000 cells (Fig. 3b). Comparing to the original mass with $k \sim N(3,1)$, the mass with $k \sim N(2,1)$ not only had a median $k=0$ but also the vast majority of cells were $k=0$ cells by 24 months, forming a dominant subpopulation with growth advantage. The final tumor with a size 10^{12} cells appeared to be homogeneous in k values. These simulations indicate how a cell was selected due to its k value when a mass grew. The presence of many cells with $k=0$ in a mass will provide a potential for unlimited growth, due to the lack of a resistance potential.

Table 1. The probability of reduction of the differentiation coefficient to below 0 or 2 by genetic alterations alone, given a frequency of 1 alteration/generation.

Generation	k_p	Prob. $k \leq 0$	Prob. $k \leq 2$
1	1.25	2×10^{-36}	
2	2.09	6×10^{-50}	2×10^{-1}
3	2.65	2×10^{-52}	7×10^{-5}
4	3.02	3×10^{-49}	1×10^{-7}
5	3.28	1×10^{-45}	4×10^{-9}
6	3.45	7×10^{-42}	1×10^{-9}
7	3.56	2×10^{-38}	1×10^{-9}
8	3.64	3×10^{-35}	3×10^{-9}
9	3.69	2×10^{-32}	7×10^{-9}
10	3.72	4×10^{-30}	2×10^{-8}
11	3.75	6×10^{-28}	6×10^{-8}
12	3.76	4×10^{-26}	2×10^{-7}
13	3.77	2×10^{-24}	4×10^{-7}
14	3.78	6×10^{-23}	8×10^{-7}
15	3.78	6×10^{-23}	2×10^{-6}

The generation specific k value for a cell is calculated according to the equation 11 and the generation number is derived from equation 5. The accumulated probability for mutations to reduce the generation specific k value to 0 or 2 is calculated according to equation 7. Since $m_i \sim N(0,0.1)$ are independent by assumption, the random variable $M = \sum_{i=1}^n m_i$ is also normally distributed, with $M \sim N(0,0.1\sqrt{n})$. This table shows the probability $P(M \leq L - k_p)$, where $L = 0$ or 2.

doi:10.1371/journal.pone.0016859.t001

This simulation describes one cascade of tumor development and evolution of a dominant subpopulation, where the differentiation coefficient in some cells is already zero (completely transformed). Evolution during the early stage of tumorigenesis is less prominent since the difference of k values is small among cells. In this case, environmental factors play an important role in the selection of cells with growth advantage. Therefore, with continuous occurrence of mutations and several cascades of evolution, a substantially transformed cell will emerge and expand as a subpopulation. The entire tumor although still heterogeneous, will become increasingly aggressive. Furthermore, as discussed below, positive and strong stimulations can promote the growth of cells with $k > 0$, creating a clinically detectable mass consisting of less aggressive cells (with substantially reduced but non-zero k values).

Calculation of cancer incidences including subclinical (dormant/transient) and clinically detectable tumors

According to the simulation of malignant transformation by genetic alterations in Table 1, the chance of getting a cell with $k = 2$ is 6×10^{-8} , suggesting that 60,000 cells would have a $k = 2$ in a human's lifetime with 10^{12} cells. We carried out simulations to test how many cells among those with $k = 2$ will escape senescence and develop into tumors under certain stimulatory conditions. All cells were already at the senescent stage (defined as $g \geq 11$), and for further simulations, we assumed an α_p value of -2/month. Hormonal stimulation could be very strong during the menstrual cycle. However, we did not include strong cyclic estrogen but instead included a slightly positive stimulation with $\beta \sim N(3,1)$ and $\beta \sim N(4,1)$ to represent chronic exposure to a weak estrogen in a

menopausal woman. Simulations were also performed for various mutational rates. Data shown in Tables 2 and 3 suggest that escape of senescence and development of a mass is positively related to the estrogenic stimulation and mutation rate with the parameter values mentioned earlier. A positive stimulation $\beta \sim N(4,1)$ and higher mutation rate (2 mutations per generation) would have 60,000 cells escaping senescence in an individual if the condition lasted for 6 months, including 87 masses with more than 10 cells, 4 masses with more than 100 cells, 3 masses with more than 1,000 cells, 2 masses with more than 10,000 cells, 2 masses with more than 100,000 cells and 1 mass with more than 1,000,000 cells (Table 2). However, if the exposure lasted for one year, the incidence of neoplasm would dramatically increase (Table 3). An individual would have 60,000 cells escaping senescence, 3,523 masses with more than 10 cells, 869 masses with more than 100 cells, 424 masses with 1,000 cells, 286 masses with more than 10,000 cells, 215 masses with more than 100,000 cells and 164 masses with more than 1 million cells. The latter masses would be large enough to be detected in the clinic. When the mutation rate was reduced to one per generation, or stimulation down to $\beta \sim N(3,1)$ or both, the incidence of tumors with varying sizes was substantially less (Tables 2 and 3). Thus, the incidence of clinical and subclinical forms of neoplasm is affected by mutation rate and exposure to chronic stimulation. If these factors can be quantified, then the incidence of cancer can be predicted for any specific individual. As a general rule, the cancer incidence either subclinical or clinical is positively correlated with an individual's age, hormonal stimulation and mutagenic exposure. This simulation is consistent with reports that subclinical forms of dormant and transient tumors could be commonly present in an individual and the incidence could be much higher if multicentric tumors are included [20–23].

An alternative potential mechanism of the initiation of cancer metastasis: destruction of inter-cellular structure resulting from massive cell death

Since a normal endometrium contains billions of cells, there are multiple clones in different stages of their lifetime providing the tissue a stable total cell number [24]. We hypothesize the existence of many basic tissues. Each of them maintains a stable cell number with a minimal number of developing clones. The maintenance of a stable number of cells in a basic tissue is thus the most fundamental phenomenon of homeostasis at the tissue level [24]. Any sudden increase in cell number (volume) of a tumor will elicit negative reaction from the microenvironment either due to the lack of factors essential for cell survival or due to physical and biochemical disruption of the environment (spatial dependence). Using equations 10 and 11, we arrived at the estimated stable cell number of 18,800 for a basic tissue in the endometrium if a stable cell number was defined as variation of cell number at $\leq 5\%$ at any time. Obviously, a basic tissue consists of many clones at different stages of their lifetime. To study the interaction between growing cells and their environment, we introduced two concepts: **cellular impact** (C_i) through which a cell exerts its effect on the basic tissue, and **ecological balance** (B) which is the direct reaction from the basic tissue. Ecological balance is a reciprocal action to all cells' impacts ($\sum C_i$) on a basic tissue.

For any individual cell, the instantaneous impact of the cell on the basic tissue at time t is proportional to its rate of growth, $\frac{dN_i}{dt}$ (equation 2), so

$$C_i = \gamma \alpha_i N_i$$

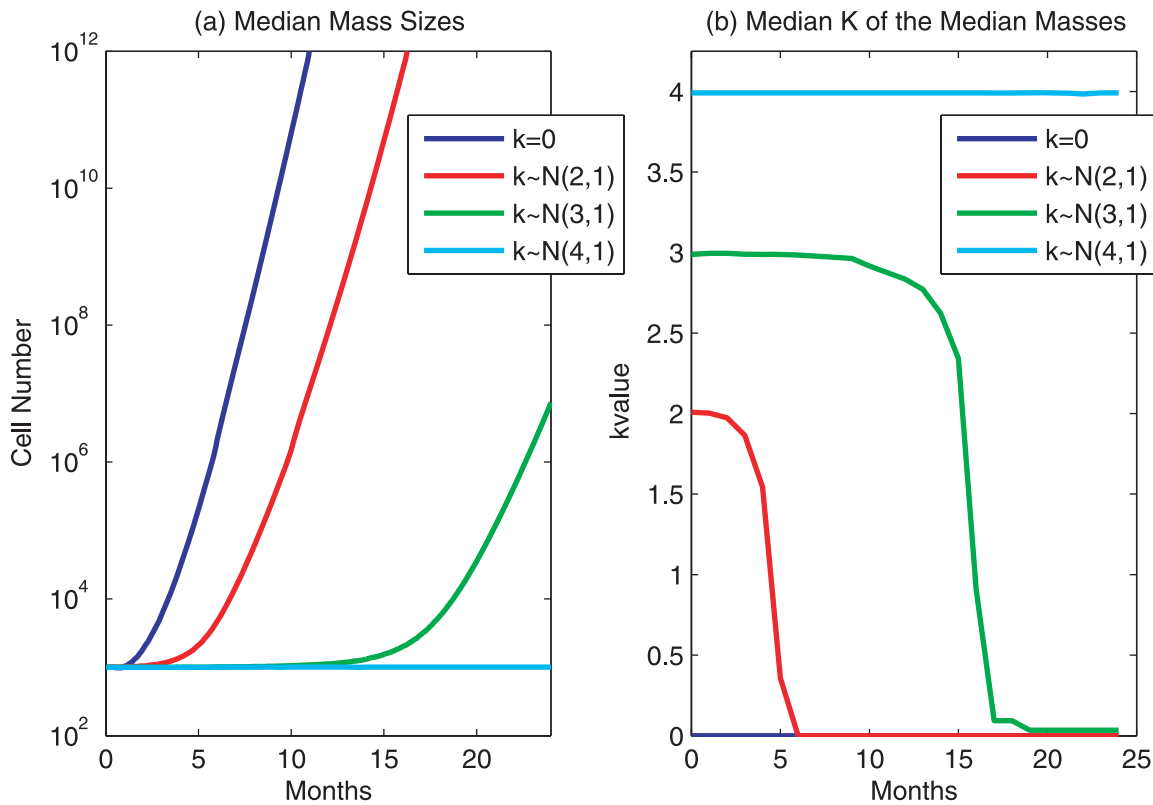


Figure 3. Simulation of tumor evolution. Tumor size change and re-distribution of intra-tumor subpopulations are simulated. **a. The growth curve of four masses with 1000 heterogeneous cells each, specified by different differentiation coefficients to indicate the extent of loss of differentiation.** Every cell has an $\alpha_p = 0$ and is assigned at $t = 0$, randomly and independently, an $\alpha(0) \sim N(0,1)$ to indicate heterogeneous initial proliferation potentials. Each mass has a k value for its cells from $k \sim N(0,0)$, $k \sim N(2,1)$, $k \sim N(3,1)$, and $k \sim N(4,1)$, respectively. All cells grow at a neutral condition $\beta \sim N(0,1)$. The tumor size is expressed as the total number of cells within a mass, with each individual cell proliferating according to the system 9. A simulation is terminated when the total number of cells exceeds 10^{12} . No mutation is considered thus the k value remains constant for each cell and its descendants. **b. Change of the values of differentiation coefficients over time in the four tumors.** The median k values over time among all the cells of one of 4 masses shown in Fig 3a are shown. The k values shown in the legend are the initial distributions of k values within each of the four masses at time $t = 0$. doi:10.1371/journal.pone.0016859.g003

for some proportionality constant γ . Summing over all N_s cells,

$$C = \sum_{i=1}^{N_s} C_i = \gamma \left(\sum_{i=1}^{N_s} \alpha_i \frac{N_i}{N_s} \right) = \gamma \alpha_\mu N_s,$$

where α_μ is the expected value of the α_i 's. In a basic tissue, there is a stable cell number, so it should hold that $\alpha_\mu \approx 0$. It follows that $C \approx 0$ in a basic tissue.

Ecological balance will not act on a specific cell as a direct reaction to its C_i value, but will instead affect all cells in the basic tissue in a stochastic fashion following a Gaussian distribution using B_μ and B_σ values. Thus,

$$B_\mu = -C_\mu = \left(\sum_{i=1}^{N_s} C_i \right) / N_s, \tag{12}$$

$$B_\sigma = C_\sigma = \left(\sum_{i=1}^{N_s} (C_\mu - C_i)^2 / N_s \right)^{1/2}, \tag{13}$$

In order to simulate a potentially negative effect from the microenvironment on a fast growing tumor and investigate the effect of differentiation coefficients on the growth, we conducted simulations of four neoplastic cells with $k = 0, 0.5, 1$ and 2 arising

in four separate basic tissues. They grew under identical conditions with $\beta_1 \sim N(1/month^2, 0.2)$ and an initial $\alpha(0) = 1/month$. The environmental reactions to tumor growth, $\beta_2 \sim N(B_\mu, B_\sigma)$ were calculated according to Equations 12 and 13, and assigned to every surviving cell every day randomly and independently. Our results showed that for the first 12 months, the cell with $k = 0$ proliferated very fast and reached the median mass size of approximately 170,000 cells by 5 months (Fig. 4a). A cell death of more than 6000 cells occurred by 5 months (Fig. 4b), which was followed by a slow increase in mass size but accompanied by increasing cell death. More than 7 million cells died while the mass size reached only 680,000 cells by 12 months (Fig. 4a and 4b), suggesting that the number of cell deaths is more than 10 times the mass size. The cell with $k = 0.5$ proliferated and reached the similar mass size by 12 months (Fig. 4a). However, significant cell death did not occur until the 9th month (Fig. 4b). The two cells with $k = 1$ and $k = 2$ grew very slowly (Fig. 4a) and did not have any substantial cell death by 12 months (Fig. 4b).

The above simulations suggest an interesting possibility. Massive cell death causes destruction in the tissue, resulting in widespread cell relocation and rupture of blood vessels. Thus, we speculated that massive cell death could lead to cell motility and tumor expansion. Simulations conducted by Enderling et al suggests a similar role of cell death for tumor expansion and invasion [19]. The simulations here showed that the faster proliferation of a more

Table 2. Simulation of the effect of mutation rate and environmental factors on the incidence of neoplasm in an individual.

Magnitude	$N(3,1),1/g$	$N(3,1),2/g$	$N(4,1),1/g$	$N(4,1),2/g$
$\geq 10^0$	60,000	60,000	60,000	60,000
$\geq 10^1$	0	0	0.76	87
$\geq 10^2$	0	0	0	4
$\geq 10^3$	0	0	0	3
$\geq 10^4$	0	0	0	2
$\geq 10^5$	0	0	0	2
$\geq 10^6$	0	0	0	1

As a continuation of Table 1, an individual was assumed to have an incidence of 60,000 cells with $k=2$ in his life time. These cells were assumed to be in senescent stages ($x_p = -2$) with slight proliferation potential ($\alpha(0)=0.2$). Different mutational rates, 1 (1m/g) or 2 (2m/g) per generation, and varying environmental stimulations, $\beta \sim N(3,1)$ or $\beta \sim N(4,1)$ were used to investigate the incidence of escape from senescence and development of different sizes of masses by each of 60,000 cells. Exposure times to chronic stimulation were 6 months every year (Table 2) or all year around (Table 3). Simulations are run for a period of 1 year. Outcome of 60,000 cells with $k=2$ under hormonal exposure for 6 months.

doi:10.1371/journal.pone.0016859.t002

Table 3. Simulation of the effect of mutation rate and environmental factors on the incidence of neoplasm in an individual.

Magnitude	$N(3,1),1/g$	$N(3,1),2/g$	$N(4,1),1/g$	$N(4,1),2/g$
$\geq 10^0$	60,000	60,000	60,000	60,000
$\geq 10^1$	0	0.48	728	3532
$\geq 10^2$	0	0.1	30	869
$\geq 10^3$	0	0.02	5	424
$\geq 10^4$	0	0.01	2	286
$\geq 10^5$	0	0	0.87	215
$\geq 10^6$	0	0	0.5	164

As a continuation of Table 1, an individual was assumed to have an incidence of 60,000 cells with $k=2$ in his life time. These cells were assumed to be in senescent stages ($x_p = -2$) with slight proliferation potential ($\alpha(0)=0.2$). Different mutational rates, 1 (1m/g) or 2 (2m/g) per generation, and varying environmental stimulations, $\beta \sim N(3,1)$ or $\beta \sim N(4,1)$ were used to investigate the incidence of escape from senescence and development of different sizes of masses by each of 60,000 cells. Exposure times to chronic stimulation were 6 months every year (Table 2) or all year around (Table 3). Simulations are run for a period of 1 year. Outcome of 60,000 cells with $k=2$ under hormonal exposure for 12 months.

doi:10.1371/journal.pone.0016859.t003

aggressive tumor with lower k cells did not necessarily lead to an overtly larger tumor but instead might result in earlier tissue invasion and metastasis. This is consistent with clinical observations that an aggressive cancer does not often present as a solitary large primary mass but a benign tumor does. As indicated by the simulations, tissue invasion and metastasis might occur in the fifth month when the mass size of $k=0$ tumor was approximately 170,000 cells.

However, the simulations above did not consider the positive interactions between cancer cells and microenvironment. Cancer cells may release some factors to enhance angiogenesis, and the surrounding tissues may adjust to accommodate the growing mass.

These positive interactions may catch up to the increase in tumor size and reach a new balance for further tumor growth. In addition, a slowly proliferating tumor may have a good chance to reach a new balance and achieve a larger tumor size without metastasis.

Survival of metastatic cells in the circulation and development of clinically detectable metastatic lesions

We now investigate the likelihood of cancer cell survival during the metastatic process and the role of the differentiation coefficient in the establishment of metastatic lesions. We assume that a cancer cell takes 5 steps to reach an ectopic (distant) site and attain

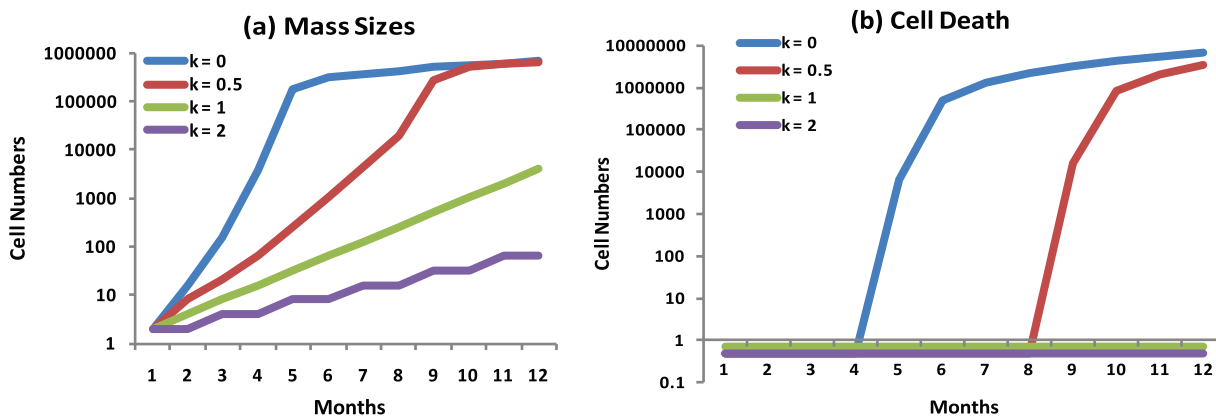


Figure 4. Cell death during tumor progression. Four cells with $k=0$, $k=0.5$, $k=1$ and $k=2/month$, respectively, are growing in an identical growth condition provided by a $\beta_1 \sim N(1/month^2, 0.2)$, a weak and positive growth stimulation, and an identical initial proliferation potential $\alpha(0)=1/month$. Environmental reaction to tumor growth is based upon $\beta_2 \sim N(B_\mu, B_\sigma)$, which affects any cell in the basic tissue in a stochastic fashion. The environmental reaction is defined as an ecological balance and is calculated according to equations 12 and 13. No mutation is included, thus the k value remains constant for each cell and its descendants in the simulations. **a.** The mass sizes arising from the four cells with different k values but identical β_1 value and initial α value. **b.** The accumulated numbers of cell death are presented for the four masses with $k=0$, 0.5, 1 and 2, respectively. The accumulated total cell death over 12 months for the cells with $k=1$ and $k=2$ are zero but presented at a value of 0.5 due to the logarithm scale of the y-axis. A $\beta_2 \sim N(B_\mu, B_\sigma)$ is calculated and assigned to each surviving cell every day.

doi:10.1371/journal.pone.0016859.g004

proliferation for the establishment of a metastatic lesion according to the main steps in the formation of a metastasis [25]. Each step is rate-limiting and modeled as negative growth stimulation. Three masses each had 10,000 cells migrating to a distant site and their k values for every one of 10,000 cells were $k=0$, $k=1$, and $k=2$, respectively (Fig. 5). The implemented β values for each rate-limiting step are: $\beta \sim N(-6/month^2, 0.2)$ representing dissemination at the primary site for 3 days, $\beta \sim N(-10/month^2, 0.2)$ representing intravasation for 2 days, $\beta \sim N(-20/month^2, 0.2)$ representing circulation for 1 day, $\beta \sim N(-10/month^2, 0.2)$ representing extravasation for 2 days, and survival at the ectopic site with $\beta \sim N(-6/month^2, 0.2)$ for 3 days, following a similar schedule in experiments [26]. Metastasis is assumed to be followed by a $\beta=0$ for 5 years without distinguishing favorable and unfavorable sites. Cells from the three masses have shown a striking difference in their survival and establishment of metastatic lesions. The survival rate at the end of the metastatic cascade (after 11 days) is similar among the three masses with more than 90% survival (Fig. 5a and 5b), consistent with experimental data showing that majority cells survive the metastatic process [26,27]. However, the mechanism for metastatic inefficiency is different among cells with different levels of malignancy. A more malignant cell (with $k=0$) was more vulnerable to the harsh conditions during metastasis and had the lower survival rate (Fig. 5b). Only 19 (0.19%) out of 10,000 cells with $k=0$ survived, consistent with *metastatic inefficiency* primarily due to subsequent cell death resulting from apoptosis [28,29]. The surviving rates for cells with $k=1$ and $k=2$ were 1.3% (132 out of 10,000) and 11.8% (1179 out of 10,000), respectively (Fig. 5b). But none of them could retain sufficient growth to develop into a clinically detectable metastatic

lesion over 5 years (Fig.5a), which is consistent with the mechanism of metastatic inefficiency due to failure of subsequent growth [26]. Only those cancer cells with $k=0$ could grow into a metastatic lesion after a successful metastasis (Fig. 5a). Cells with $k \geq 1$ remained dormant without further mutations and environmental stimulations. Thus, establishment of clinically significant ectopic lesions correlates positively with the number of cells with extremely low k value in a primary tumor and the fast growth of the primary tumor (Fig. 4), but negatively with the extent of harsh conditions in the metastatic process (Fig.5). This result is consistent with the well-established role of metastatic lesions to predict patient outcome. But our model provides an approach to quantitatively and prospectively predict the chance of a metastatic lesion through analysis of the primary tumor, and may be particularly useful to predict the existence of micrometastasis. For cells with $k > 0$ to grow in ectopic sites, a positive β is required. The value of β determines whether an ectopic site is favorable for the development of metastatic lesions. Thus, the long-held hypothesis of *seed and soil* for pathogenesis of metastasis [25,29] is quantitatively modeled here.

These results are also consistent with clinical observations that metastatic endometrial cells are not cancer cells in endometriosis. Endometrial cells undergo a fast proliferation during the early menstrual cycle due to estrogenic stimulation and massive cell death occurs late in the cycle due to withdrawal of hormones. Some endometrial cells are transported through fallopian tubes and implanted on the surface of the pelvic cavity. These endometrial cells survive the metastatic process and the condition in ectopic sites due to their high differentiation coefficient. As with cells in the primary site, ectopic endometrial cells only proliferate

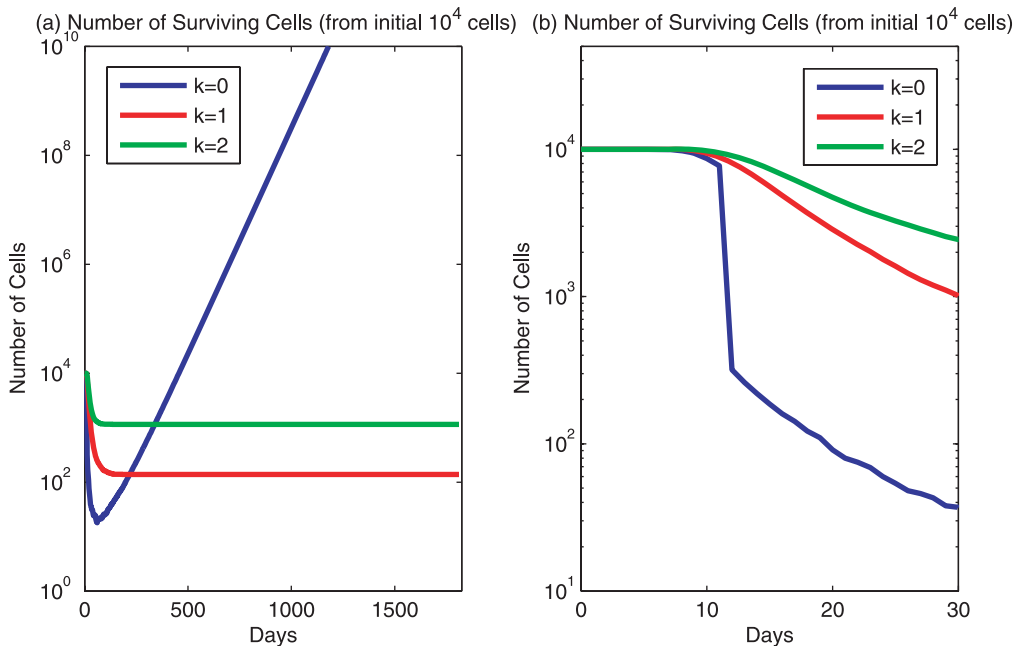


Figure 5. Modeling of cancer cell survival. The metastatic process and the development of clinically detectable metastatic lesions are simulated. The number of cells surviving over the course of a metastatic process is shown. Three masses of 10,000 cells each attempt migration to a distant site, with all cells within a mass possessing either $k=0$, $k=1$, and $k=2$, respectively. Every cell has an initial $\alpha(0) \sim N(0,1)$ and must pass 5 steps, which are assumed to be negative for any cell to grow. Every step has its specific β value: dissemination at the primary site $\beta \sim N(-6/month^2, 0.2)$ for 3 days, intravasation $\beta \sim N(-10/month^2, 0.2)$ for 2 days, circulation $\beta \sim N(-20/month^2, 0.2)$ for 1 day, extravasation $\beta \sim N(-10/month^2, 0.2)$ for 2 days, and survival at the ectopic site $\beta \sim N(-6/month^2, 0.2)$ for 3 days. **a.** Cancer cell growth is followed for 5 years to show clinically detectable lesions. **b.** The same cancer cells are followed for only 30 days in order to show the number of surviving cells when they reached distant sites ($t=11$ days). The surviving cell number is the summation of all cells whose growth is governed by system 9. No further mutation is considered thus the k value remains constant for each cell and its descendants in the simulations. doi:10.1371/journal.pone.0016859.g005

in response to high estrogen. Large β is derived from high estrogen at the proliferative phase of a menstrual cycle, similar to those modeled at Fig. 2b. The ectopic endometrial tissues become atrophic with prolonged progestin treatment, pregnancy and at menopause. Thus, the occurrence of endometriosis is due to 1) successful metastasis due to mechanical transportation of endometrial tissues into ectopic locations; 2) high survivability of cells due to their high differentiation coefficient; and 3) periodic proliferation stimulated by strong and cyclic hormones. Most important of all, these ectopic endometrial cells have high k values and are not cancer cells.

Discussion

It has been the hope of every cancer researcher that complex cancer biology can be rationalized based on a small number of underlying principles which govern the myriad cancer genotypes and phenotypes [3,30]. The exceptional elucidation of the general principles of cancer formulated by Hanahan and Weinberg, here referred to as the W theory [6], coupled with efforts to elucidate the mechanisms leading to carcinogenesis exemplified by the seminal work of Vogelstein and colleagues, here referred to as the V theory [4,5], combine to provide a solid heuristic foundation, referred to here as V/W. Our model, using simple mathematical equations featuring a small number of variables, is an attempt to quantitatively and prospectively interpret V/W. It is also an attempt to develop a classical *insight and theory* approach wherein principles and equations are derived from fundamental concepts and deductions as opposed to a direct a priori analysis of data followed by an a posteriori explanatory model.

Assuming in our model that cells targeted for malignant transformation include well-differentiated cells, malignant transformation can be considered to be a process of de-differentiation [17], which can be quantitatively described by a differentiation coefficient. As illustrated at Fig. 1, numerous mutations occur in a tissue with billions of cells constantly turning-over. Most non-stem uterine epithelial cells have a lifespan limited to days or months. The cumulative effect of mutations is neutral and not transformative in vast majority of normal cells since the mutational effect follows a Gaussian distribution as proposed by our model. One out of billions of cells acquires a mutation, or several mutations, which result in de-differentiation, which is captured in a differentiation coefficient, k in our model (equation 7). A reduced k value results in a lower resistance potential (equation 6), which will allow the cell to multiply in a subnormal level of hormonal (or other) stimulation (system 9) and create a population of cells well beyond homeostasis that might initially be hyperplastic or benign. Additional mutations can further reduce cells' k values yielding selective fitness and a dominant sub-population due to their lower k values (Fig.3). A lower k value corresponds to further de-differentiation thus further malignant transformation. Again, some cells in this population will emerge with even lower k values and continue the cascade of malignant transformation and tumor evolution. This mathematical expression of multistage oncogenesis is strikingly complementary to the V theory [4,5]. Additionally, as simulated in the Fig. 4, rapid growth of an increasingly malignant cell population will be accompanied by massive cell death thus creating a potential mechanism of metastasis that does not require but facilitate further mutations. This result provides a quantitative interpretation of Jones' genomic analysis [5] opposing the long held paradigm that the appearance of metastatic cell populations is driven by additional mutations [4,5].

The V theory of tumorigenesis posits the fundamental role of genetic mutations in malignant transformation as the pathway for

tumor progression. While it is likely that only a small number of mutations are driver mutations and the rest are passengers, our model accounts for the contribution of every potential mutation in a normal distribution and, from this, the collective roles played by all mutations can be quantified. The precise contribution of a specific mutation has to be analyzed in patient samples [12]. Sequencing of many tumor genomes is increasingly being realized with the initiation of Cancer Genome Atlas and Cancer Genome Anatomy project.

Tumor evolution is described by many investigators analogously to Darwinism with tumors being a clonally-derived cell populations and progression the result of some cells acquiring advantages through mutations [31,32]. The descriptive theories are complemented by our quantitative analysis that has a potential to interpret a wide range of different cancer phenotypes. Use of our model either in population level or single cell level may prove useful in the reconciliation of some seemingly contradicting observations. Advancing age is the most potent of all carcinogens [33]. On the other hand, cellular senescence (ageing) is as potent as apoptosis in suppressing spontaneous tumorigenesis [34]. In our model the probability of malignant transformation increases significantly as generations progress at the single cell level. However, senescence significantly increases when generation progresses (equation 3). These factors play against each other and the balance will be shifted to favor transformation if environmental stimulation increases to escape senescence by providing cell proliferation beyond intrinsic potential. At the tissue level, the probability of emergence of a malignant cell will increase with age since an older individual will have more cell cycles in their lifetime and, consequently, increased cancer incidence.

Our model uses cellular generations as a biologic clock which is a fundamental variable in determining cellular proliferation and differentiation. While cell division is intuitively simple and clear from a morphological perspective, a mathematical expression is necessary to provide a usable description of propagation of an individual cell during a clone lifetime. Cell divisions are determined by proliferation potential that depend on proliferation history and specific factors, such as mutations and environmental cues. This approach allows us to follow the time course of a cell during a clone lifetime without a need for counting cell divisions experimentally, therefore measuring intra-tumor heterogeneity at any time quantitatively. A proposed generation limitation is a quantitative interpretation of replicative senescence.

Our model simultaneously describes individual cells as well as cell populations (or tumors), thus preserving population heterogeneity. This approach captures qualitative change in individual cells as the critical event for tumor progression (cascade of tumor evolution). This additive approach to summarizing individual cells may allow us to prospectively describe incidences of dormant/transient cancers (those cancers either not easily diagnosed clinically or not diagnosed at all [20,23]). In the future, detailed comparison between lethal cancer and non-disease cancer will be a much more informative way in which to identify truly causal genetic alterations. In a normal physiologic scenario, growth of a clone or an individual cell of the clone will have minimal effect on the environment. However, growth of a malignant mass is fast and can be extensive, exceeding normal physiological limits. Thus, the impact of the mass on its surroundings will be significant and, conversely, the reaction of the surrounding environment will also be dramatic (equation 12). This will cause massive cell death (Fig. 4), here identified as a potential mechanism of the initiation of metastasis.

Our model is best described as a classic theoretical model of *insight and theory* approach and the coherence of the model is fully

embedded in the mathematical equations. It is a novel approach and still at an early stage of development. Numerous tissue- and cell-specific variables have to be quantitatively determined before such a model can be experimentally applied to a wide range of cancer phenomena. Modeling of environment effects in this model also will require modification over time. Fortunately, there are many excellent models regarding the role of environment factors in oncogenesis and tumor progression [35–39]. Our hope is that we may be able to integrate our model with those existing models. Such a theoretical approach should be complimentary to empirical and bioinformatic approaches to cancer research in a way that data are used iteratively to test and update equations and laws.

Taken together, our modeling of malignant transformation and tumor progression, based on a few assumptions and a few mathematical equations, is capable of comprehensively describing major cancer phenotypes and has suggested some interesting features for neoplasm: 1) incidence of neoplasm is much higher than what is diagnosed in clinics due to the frequent existence of dormant/transient neoplasm; 2) neoplastic cells are heterogeneous in a tumor and the aggressive cells (with low differentiation coefficients) have a growth advantage that allows them to dominate the tumor during its evolution; 3) rapid proliferation can result in cell death and lead to cell dislocation and metastasis; 4) more aggressive cancer cells have less chance to survive metastasis but are more capable of growing into a clinically significant lesion. Therefore, an aggressive tumor, defined by the presence of cancer cells with low differentiation coefficients, will have a poor patient outcome since the significant number of $k \approx 0$ cells will ensure fast growth in the primary and ectopic sites in most environmental conditions.

Methods

The Model

In this paper, we discuss a fundamental model for cancer that has shown the capability to systematically capture many varied cancer phenotypes. The quantitative interpretation of oncogenesis is preceded by a comprehensive description of normal tissue regeneration in uterine epithelium, based on the assumption that every one of billions of normal cells follows an inherent program for proliferation and differentiation through a cellular interpretation of chronological time. The landscape of physiological tissue regeneration provides a picture of normal cell heterogeneity and dynamics, and serves as the reference for tumor development resulting from genetic insults and environmental stimulations. It is a single-cell based model, where the evolution of each tumor is described in terms of the evolutionary dynamics of its constituent cells. The quantitative basis of the model is a quantification of a cell's differentiation status, a measurement of how “normal” a cell is, and how this status varies due to different mutations. The cellular dynamics are defined on a per cell basis by the system of equations:

$$n(t) = \int_{t_n}^t \alpha(s) ds, \quad \frac{d\alpha}{dt} = r(t) + \beta(t), \quad r(t) = k(t) \cdot (\alpha_p(t) - \alpha(t)),$$

$$k(t) = k_p(t) + \sum_{i=1}^n m_i,$$

where $\alpha(t)$ is a cell's proliferation potential at time t and $k(t)$ is its differentiation coefficient. The parameters m_i are the mutational effects on the differentiation coefficient, and are distributed according to a Gaussian distribution with mean μ and standard deviation σ , $m_i \sim N(\mu, \sigma)$. The parameter $\beta(t)$ is the environmental

effect and is considered as an overall effect over the course of one day, and is also chosen from a Gaussian distribution. Each cell has associated with it certain programmed levels of proliferation potentials, $\alpha_p(t) = (1/L') \cdot (L - g(t)) \cdot g(t)$, and differentiation coefficients, $k_p(t) = K_{max} [1 - e^{(-K' \cdot g(t))}]$, which are dependent on the maturity of the cell, measured by its generation number,

$g(t) = ceiling(\int_0^t |\alpha(s)| ds)$. The resistance potential ($r(t)$) is a restorative force, acting to restore α back to its programmed level. A measure of the cell's status toward either dividing or dying is provided by $n(t)$, where the cell was born at time t_n . This system of equations is valid over the lifetime of the cell, that is, from the time it is born (t_n) until it either dies ($n(t) = -1$) or divides ($n(t) = 1$). In the case that the cell divides, the two daughter cells inherit α , k , and g from the parental cell. Each cell then proliferates independently of the other according to this system of equations. The developmental process continues with cell-specific β 's, and the mutations accumulated by each daughter cell are independent of those accumulated by the other daughter cell.

In all of the simulations presented in this manuscript, we consider the model case where $L = 10$, $L' = 7$, $K_{max} = 3.78$, and $K' = 0.4$, so that the programmed proliferation potentials and differentiation coefficients are given by $\alpha_p(t) = (10 - g(t)) \cdot g(t) / 7$ and $k_p(t) = 3.78 \cdot [1 - e^{(-0.4g(t))}]$ respectively. In all cases considered in this manuscript, we arbitrarily take $m_i \sim N(0, 0.1)$. However, the frequency of these mutations and the distributions for the environmental effects vary among the experiments considered.

Each simulation is performed using the computer programming language Fortran 90, with the resulting graphics generated with Matlab. The time-step for each simulation is $dt = 1/30$. Over the course of one time-step, the system can be solved explicitly, with an algorithm outlining this process provided as Supporting Information S1. At time t_n , a β is chosen for each cell, and α , n , and g are advanced over dt . If the cell advances in generation number ($g_{new} > g_{old}$), the cell gains a number of mutations, dependent on the experiment. If $n \geq 1$, the cell has proliferated, with the ($2^{\lfloor n \rfloor}$) daughter cells inheriting the parent's α and k . If $n \leq -1$, then the cell is considered dead and is removed from the simulation.

The clone lifetime of uterine epithelial cells

In this section, we consider two experiments. The first experiment, illustrated by Fig. 2a, is an analysis of the clone size in uterine epithelium when there is little or no estrogen exposure. A β value is assigned to an existing cell every day, randomly and independently. Four different trials are considered: $\beta = 0$, $\beta \sim N(0, 1)$, $\beta \sim N(0, 2)$, and $\beta \sim N(1, 1)$. 101 simulations were performed for each trial, with Fig. 2a illustrating the median curves for each trial, and the maximal, 75th percentile, the median, 25th percentile, and the minimal sizes of clones are presented in Fig. S1.

The second experiment, illustrated by Fig. 2b, is an analysis of the clone size in uterine epithelium with the presence of monophasic contraceptives. The exposures to large concentrations of estrogen are represented here by strong and positive β 's. Three trials are considered: $\beta \sim N(60, 6)$, $\beta \sim N(90, 9)$, and $\beta \sim N(120, 12)$. These environmental effects are applied for 21 days before being removed. As in the previous figures, 101 simulations were performed for each trial and the median clone sizes are presented. All experiments are performed in the absence of mutations. The maximal, 75th percentile, the median, 25th percentile, and the minimal sizes of clones are presented in Fig. S2.

Malignant transformation during dynamic epithelial cell development

According to this model, the differentiation coefficients are altered solely due to mutations. In this section, we consider the accumulated probabilities of a cell to attain a differentiation coefficient of $k \leq 0$ or $k \leq 2$ after n generations. We assume that one mutation occurs per generation, with each mutation being random and independent, with $m_i \sim N(0,0.1)$. Since the mutations are independent by assumption, the random variable $M = \sum_{i=1}^n m_i$ is also normally distributed, with $M \sim N(0,0.1\sqrt{n})$. Table 1 shows the probability $P(M \leq L - k_p)$, where $L = 0$ or 2 .

Tumor evolution: growth of a heterogeneous mass consisting of cells with various differentiation coefficients

In this section, the evolution of a mass comprising cells with various differentiation coefficients is explored. Four separate trials are considered. Each trial begins with 1000 cells that have reached maturity, that is, $\alpha_p = 0$. Each cell is assigned a proliferation potential with $\alpha(0) \sim N(0,1)$. Every cell was assigned a k -value from either $N(0,0)$, $(2,1)$, $(3,1)$ and $(4,1)$, depending on the trial. Only nonnegative k values are considered in this paper, so if $k < 0$, we set $k = 0$. No mutations were considered, and therefore each cell and its descendants retained their initial k -value. For each cell, the environmental effects were chosen with a neutral mean, $\beta \sim N(0,1)$. The simulations were performed until either a simulation period of 2 years had passed or the total cell number had reached 10^{12} . Fig. 3a represents the median mass sizes of 101 simulations among each of the four trials, while Fig. 3b illustrates the evolution of the median k -values of the cells within the median masses of each trial. The maximal, 75th percentile, the median, 25th percentile, and the minimal values of mass sizes and k values are presented in Fig S3 and S4, respectively.

Calculation of cancer incidences including subclinical (transient and dormant) and clinically detectable tumors

In Table 1 it is shown that, for the test case, the probability of a specific cell reaching $k \leq 2$ due solely to mutations by the time it reaches programmed senescence, that is $\alpha_p < 0$ after 11 generations, is 6×10^{-8} . Assuming that there are roughly 10^{10} epithelial cells at any time, each with a lifespan of 1 year, a human with a lifetime of 100 years would have 10^{12} epithelial cells throughout their lifetime, with roughly 60,000 cells having $k \leq 2$ due solely to mutations. In this section, we analyze the evolution of these 60,000 cells to calculate empirical probabilities of having masses of various orders of magnitude during a lifetime. Two separate experiments are considered, each with four separate trials. In both experiments, we begin with 60,000 cells, each with $k = 2$. The cells begin at the senescent stage, with $\alpha_p = -2$. However, their actual proliferative potential is $\alpha(0) = 0.2$. In the first experiment, there is an environmental effect present for six months every year, while in the second experiment, the environmental effect is present year-round. The four trials for each experiment have either one or two mutations per generation, and an environmental effect following $\beta \sim N(3,1)$ or $\beta \sim N(4,1)$. For each trial, 101 simulations are performed, and empirical probabilities are calculated from the data generated. The mean number of tumors of various sizes that can be expected among the 60,000 cells are illustrated when the probability of having at least one tumor of each magnitude was one. In the case that a tumor cannot be probabilistically guaranteed, the probability of having at least one tumor of each magnitude is illustrated.

An alternative potential mechanism of the initiation of cancer metastasis: destruction of intercellular structure resulting from massive cell death

In this section, the reaction of the microenvironment surrounding the tumor is included into the model as an additional environmental effect. The impact of a cell on its environment (C_i) is proportional to its rate of growth $\frac{dN_i}{dt}$, so $C_i = \gamma \alpha_i(t) N_i$. The total environmental impact of a tumor is then

$$C = \sum_{i=1}^{N_s} C_i = \gamma \left(\sum_{i=1}^{N_s} \alpha_i \frac{N_i}{N_s} \right) = \gamma \alpha_\mu N_s,$$

where α_μ is the expected value of the α_i 's. The environmental response to this cumulative force acts on each cell in a stochastic fashion, with mean $B_\mu = -C_\mu = (\sum_{i=1}^{N_s} C_i) / N_s$ and standard deviation $B_\sigma = C_\sigma = (\sum_{i=1}^{N_s} (C_i - C_\mu)^2 / N_s)^{1/2}$. Fig. 4 illustrates the outcomes of four trials. Each trial begins with one cell. A weak growth stimulation following $\beta_1 \sim N(1,0.2)$ is chosen, with this value being the initial environmental effect for all four trials. Their initial proliferation potentials are all identically $\alpha(0) = 1$. However, their initial differentiation coefficients differ, with either $k = 0, 0.5, 1$, or 2 . No mutations are considered, so these k -values remain constant among all descendant cells. At each time step, a new growth stimulation $\beta_1 \sim N(1,0.2)$ and environmental reaction $\beta_2 \sim N(B_\mu, B_\sigma)$ are calculated. One hundred and one simulations were performed for each trial, and Fig. 4 illustrates the median curves for the mass sizes and the accumulated number of cell deaths occurring for each trial.

Survival of metastatic cells in the circulation and development of clinically detectable metastatic lesions

In this section, the likelihood of cancer cell survival during the metastatic process is explored, along with the role of the differentiation coefficient in the establishment of a metastatic lesion. The cancer cells are assumed to take five steps to reach an ectopic site and establish a metastatic lesion. The five steps are represented by negative growth stimulations. Three trials are considered. Each trial begins with 10,000 cells. The k -values of all 10,000 cells are either $k = 0, 1$, or 2 , depending on the trial. No mutations were included, so all descendants retained the same k -value. Their initial proliferation potentials are chosen from $\alpha(0) \sim N(0,1)$. Each cell underwent the following environmental effects over the course of the five steps: $\beta \sim N(-6,0.2)$ for 3 days representing dissemination at the primary site, $\beta \sim N(-10,0.2)$ for 2 days representing intravasation, $\beta \sim N(-20,0.2)$ for 1 day representing circulation, $\beta \sim N(-10,0.2)$ for 2 days representing extravasation, and $\beta \sim N(-6,0.2)$ for 3 days representing survival at the ectopic site. This was followed by an environmental effect of $\beta = 0$ for 5 years. 101 simulations were performed for each, with the median curves for the number of surviving cells illustrated in Fig. 5. The maximal, 75th percentile, the median, 25th percentile, and the minimal values of mass sizes are presented in Fig S5.

Supporting Information

Figure S1 A clone lifetime of uterine epithelial cells during menopause, illustrated by change of cell number of the clone over time. 101 simulations are performed for all cases, with the following trajectories presented here: minimum, 25th percentile, median, 75th percentile, and maximum. (a)

Results for $\beta=0$. (b) Results for $\beta \sim N(0,1)$. (c) Results for $\beta \sim N(0,2)$. (d) Results for $\beta \sim N(1,1)$. (e) Results for $\beta \sim N(2,1)$. (EPS)

Figure S2 A clone lifetime of uterine epithelial cells with oral contraceptives, illustrated by change of cell number of the clone over time. 101 simulations are performed for all cases, with the following trajectories presented here: minimum, 25th percentile, median, 75th percentile, and maximum. (a) Results for $\beta \sim N(60,6)$. (b) Results for $\beta \sim N(90,9)$. (c) Results for $\beta \sim N(120,12)$. (EPS)

Figure S3 Simulations of tumor evolution: size change of intra-tumor subpopulations of tumors with heterogeneous cells. The growth curves of four masses with 1000 heterogeneous cells each, specified by different differentiation coefficients to indicate the loss of differentiation. 101 simulations are performed for all cases, with the following trajectories presented here: minimum, 25th percentile, median, 75th percentile, and maximum. (a) Results for $k=0$. (b) Results for $k \sim N(2,1)$. (c) Results for $k \sim N(3,1)$. (d) Results for $k \sim N(4,1)$. (EPS)

Figure S4 Simulations of tumor evolution: redistribution of k-values among intra-tumor subpopulations of tumors with heterogeneous cells. Change of the median value of differentiation coefficients with respect to time for the

masses presented in Figure 3. (a) Results for $k=0$. (b) Results for $k \sim N(2,1)$. (c) Results for $k \sim N(3,1)$. (d) Results for $k \sim N(4,1)$. (EPS)

Figure S5 Modeling of cancer cell survival during metastasis and the establishment of metastatic lesions. 101 simulations are performed for all cases, with the following trajectories presented here: minimum, 25th percentile, median, 75th percentile, and maximum. (a) Results for $k=0$. (b) Results for $k=1$. (c) Results for $k=2$. (EPS)

Supporting Information S1 Computational Algorithm and Additional Figures for “Quantitative Interpretation of a Genetic Model of Carcinogenesis Using Computer Simulations”.

(PDF)

Acknowledgments

The authors would like to thank Dr. Heiko Enderling for his critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: DD. Performed the experiments: BB XW. Analyzed the data: DD. Contributed reagents/materials/analysis tools: BB. Wrote the paper: DD CH YL.

References

- Sporn M (1996) The war on cancer. *Lancet* 347: 1377–1381.
- American Cancer Society (2010) Cancer Facts and Figures.
- Weinberg R (2007) *The Biology of Cancer*. Garland Science.
- Fearon E, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61: 759–767.
- Jones S, Chen W, Parmigiani G, Diehl F, Beerenwinkel N, et al. (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* 105: 4283–4288.
- Hanahan D, Weinberg R (2000) The hallmarks of cancer. *Cell* 100: 57–70.
- Gargett C, Masuda H (2010) Adult stem cells in the endometrium. *Mol Hum Reprod* 16: 818–834.
- Spencer S, Berryman M, Garcia J, Abbott D (2004) An ordinary differential equation model for the multistep transformation to cancer. *J Theor Biol* 231: 515–524.
- Winsor C (1932) The gompertz curve as a growth curve. *Proc Natl Acad Sci USA* 18: 1–8.
- Laird A (1969) Dynamics of growth in tumors and in normal organisms. *Nat Cancer Inst Monogr* 30: 15–28.
- Skipper H, Schabel F, Lloyd H (1979) Dose-response and tumor cell repopulation rate in chemotherapeutic trials. In: *Advances in Cancer Chemotherapy*, Vol.1, Marcel Dekker. pp 205–253.
- Zhang G, Beck B, Luo W, Wu F, Kingsmore S, et al. (2011) Development of a phylogenetic tree model to investigate the role of genetic mutations in endometrial tumors. *Oncology Reports*, in press.
- Smith J, Hayflick L (1974) Variation in the life-span of clones derived from human diploid cell strains. *J Cell Biol* 62: 48–53.
- Smith J, Whitney R (1980) Intraclonal variation in proliferative potential of human diploid fibroblasts: stochastic mechanism for cellular aging. *Science* 207: 82–84.
- Blackburn E (2000) Telomere states and cell fates. *Nature* 408: 53–56.
- Blanpain C, Horsley V, Fuchs E (2007) Epithelial stem cells: turning over new leaves. *Cell* 128: 445–458.
- Riester M, Stephan-Otto A, Downey R, Singer S, Michor F (2010) A differentiation-based phylogeny of cancer subtypes. *PLoS Comput Biol* 6: e1000777.
- Bru A, Casero D (2006) The effect of pressure on the growth of tumour cell colonies. *J Theor Biol* 243: 171–180.
- Enderling H, Anderson A, Chaplain M, Beheshti A, Hlatky L, et al. (2009) Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. *Cancer Res* 69: 8814–8821.
- Folkman J, Kalluri R (2004) Cancer without disease. *Nature* 427: 787.
- Black W, Welch H (1993) Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* 328: 1237–1243.
- Sweeney E (1995) Dormant cells in columnar cell carcinoma of the thyroid. *Hum Pathol* 26: 691–693.
- Nielsen M, Thomsen J, Primdahl S, Dyrborg U, Andersen J (1987) Breast cancer and atypia among young and middle-aged women: a study of 110 medicolegal autopsies. *Br J Cancer* 56: 814–819.
- Michor F, Iwasa Y, Nowak M (2004) Dynamics of cancer progression. *Nat Rev Cancer* 4: 197–205.
- Fidler I (2003) The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nat Rev Cancer* 3: 453–458.
- Luzzi K, MacDonald I, Schmidt E, Kerkvliet N, Morris V, et al. (1998) Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* 153: 865–873.
- Chambers A, Groom A, MacDonald I (2002) Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* 2: 563–572.
- Wong C, Lee A, Shientag L, Yu J, Dong Y, et al. (2001) Apoptosis: and early event in metastatic inefficiency. *Cancer Res* 61: 333–338.
- Paget S (1889) The distribution of secondary growths in cancer of the breast. *Lancet* 1: 571–573.
- Hahn W, Weinberg R (2002) Modelling the molecular circuitry of cancer. *Nat Rev Cancer* 2: 331–341.
- Merlo L, Pepper J, Reid B, Maley C (2006) Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 6: 924–935.
- Nowell P (1976) The clonal evolution of tumor cell populations. *Science* 194: 23–28.
- DePinho R (2000) The age of cancer. *Nature* 408: 248–254.
- Deng Y, Chan S, Chang S (2008) Telomere dysfunction and tumor suppression: the senescence connection. *Nat Rev Cancer* 8: 450–458.
- Byrne H, Drasco D (2009) Individual-based and continuum models of growing cell populations: a comparison. *J Math Biol* 58: 657–687.
- Byrne H (2010) Dissecting cancer through mathematics: from the cell to the animal model. *Nat Rev Cancer* 10: 221–230.
- Gatenby R, Vincent T (2003) An evolutionary model of carcinogenesis. *Cancer Res* 63: 6212–6220.
- Gatenby R, Gillies R (2008) A microenvironmental model of carcinogenesis. *Nat Rev Cancer* 8: 56–61.
- Anderson A, Quaranta V (2008) Integrative mathematical oncology. *Nat Rev Cancer* 8: 227–234.