

# Stochastic dynamics of metastasis formation

Franziska Michor<sup>a,b,\*</sup>, Martin A. Nowak<sup>a</sup>, Yoh Iwasa<sup>c</sup>

<sup>a</sup>Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

<sup>b</sup>Harvard Society of Fellows, 78 Mount Auburn Street, Cambridge, MA 02138, USA

<sup>c</sup>Department of Biology, Kyushu University, Fukuoka 812-8581, Japan

Received 22 June 2005; received in revised form 17 October 2005; accepted 18 October 2005

Available online 15 December 2005

## Abstract

Tumor metastasis accounts for the majority of deaths in cancer patients. The metastatic behavior of cancer cells is promoted by mutations in many genes, including activation of oncogenes such as RAS and MYC. Here, we develop a mathematical framework to analyse the dynamics of mutations enabling cells to metastasize. We consider situations in which one mutation is necessary to confer metastatic ability to the cell. We study different population sizes of the main tumor and different somatic fitness values of metastatic cells. We compare mutations that are positively selected in the main tumor with those that are neutral or negatively selected, but faster at forming metastases. We study whether metastatic potential is the property of all (or the majority of) cells in the main tumor or only the property of a small subset. Our theory shows how to calculate the expected number of metastases that are formed by a tumor.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** Metastasis; Tumorigenesis; Mathematical model; Stochastic dynamics

## 1. Introduction

Metastases arise when cancer cells leave the primary tumor site and form new tumors in another organ or elsewhere. When a tumor is detected at an early stage, before it has spread, it can often be treated successfully by surgery, irradiation or chemotherapy, and the patient can be cured. When a cancer is diagnosed once it has metastasized, however, treatment outlooks are much dimmer. Furthermore, metastases are detected in many patients after the removal of their primary tumors, even if the tumors had apparently not yet spread at the time of the initial diagnosis. These metastases can show an organ-specific pattern of spread. For example, breast and prostate cancers often metastasize to bone, and might occur years or even decades after apparently successful primary treatment.

Metastasis formation is a complex, multistep process (Weiss, 2000; Chambers et al., 2000, 2001, 2002) (Fig. 1). Initially, the in situ tumor is surrounded by an intact basement membrane. Invasion of the tumor border requires changes in adhesion, initiation of motility and proteolysis of the extracellular matrix. These changes result in the shedding of cancer cells into the circulation, either directly or via the lymphatic system. Subsequently, the migratory cells must survive in the circulation, be arrested in a distant organ and extravasate from the capillary. These steps have been observed to occur quite efficiently by in vivo microscopy of melanoma cells (Chambers et al., 2000, 2001, 2002). The next and last phase of metastasis, collectively known as metastatic colonization, is subdivided into survival of cells after extravasation, initial growth after extravasation and persistence of growth. Each of these stages of metastatic colonization occurs inefficiently, and the rate of invasion and metastatic colonization predicts the overall metastatic ability.

Despite the obvious importance of metastasis, the process remains incompletely characterized at the molecular and biochemical levels (Fidler, 2002). Transfection studies indicate that many genes can induce metastatic

\*Corresponding author. Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA 02138, USA.

Tel.: +1 617 496 4698; fax: +1 617 496 4629.

E-mail address: [michor@fas.harvard.edu](mailto:michor@fas.harvard.edu) (F. Michor).

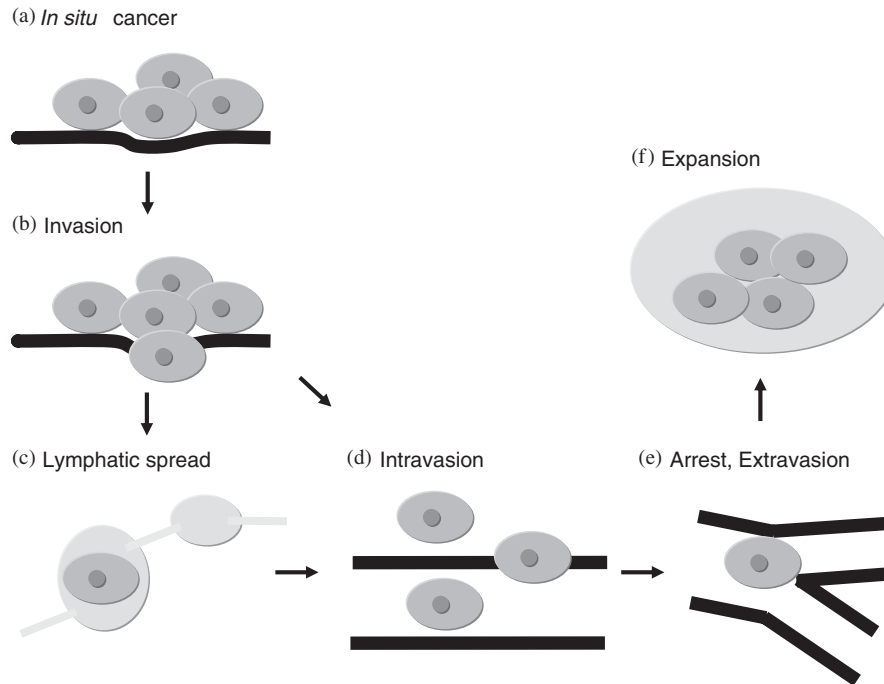


Fig. 1. The metastatic process begins with (a) an *in situ* cancer surrounded by an intact basement membrane. (b) Invasion requires changes in the cell–cell and cell–extracellular-matrix adherence and motility. Metastasizing cells can (c) spread via the lymphatics, or (d) directly enter the circulation. After survival in the circulation, arrest and extravasation (e), single metastatic cells colonize the distant site (f), giving rise to dormant cells, occult micrometastases and/or progressively growing, angiogenic metastases.

ability. Transfection of early passage embryo cells by the oncogene RAS results in fully malignant cells with metastatic potential (Pozzatti et al., 1986), and expression of the MYC oncogene confers metastatic potential to fibroblast tumors (Wyllie et al., 1987). Expression of activated ERBB2/NEU oncogene is sufficient to induce experimental metastasis (Yu and Hung, 1991).

The identification of RB as a tumor suppressor gene led to the hypothesis that metastasis might also involve loss of gene functions that maintain the normal state of a cell. In 1988, the first metastasis suppressor gene, NM23, was identified and many other metastasis suppressor genes have been discovered in the meanwhile (Steege et al., 1988). The dynamics of metastasis suppressor gene inactivation will not be discussed in this paper, but is the topic of ongoing investigation (Michor et al., forthcoming).

A long-standing question in metastasis research concerns the time in tumor progression at which cancer cells acquire metastatic identity (Bernards and Weinberg, 2002; Veer and Weigelt, 2003). According to the prevailing reasoning, tumor progression initially gives rise to cells that form tumor masses of substantial size. Subsequently, individual cells in these large cell populations acquire yet more mutations that enable them to metastasize.

A different mechanistic model proposes that the tendency to metastasize is largely determined by mutations acquired relatively early in tumorigenesis (Bernards and Weinberg, 2002). This hypothesis is based on findings that in human cancers, expression profiles of the whole primary

tumor can predict disease outcome of cancer patients (Sorlie et al., 2001; van de Vijver et al., 2002; Veer et al., 2002; Ramaswamy et al., 2003). A poor-prognosis signature is strongly predictive for the development of distant metastases, in contrast to the ‘good-prognosis’ signature.

Mathematical models of metastasis have provided insights into the prognostics and treatment outcomes. Bosl et al. offer a multivariate analysis of prognostic variables in patients with metastatic testicular cancer (Bosl et al., 1983). A competition model describes tumor–normal cell interaction with chemotherapy and parameter conditions needed to prevent relapse following attempts to remove the tumor or tumor metastases (Panetta, 1996). Another mathematical model determines the frequency of first metastatic events in breast cancer (Thames et al., 1999). Pescarmona et al. develop a nonlinear model of cancer growth and metastasis (Pescarmona et al., 1999). A mathematical model to simulate the spatiotemporal evolution of neoplasias leads to four different outcomes; indefinite growth, metastasis, latency, and complete regression (Delsanto et al., 2000). Mathematical models are also used to examine the role of genetic instability in angiogenesis and metastasis (Wodarz and Krakauer, 2001). A mathematical model of axillary lymph node involvement in breast cancer metastases shows that surgical sampling on the basis of lymph node size might have good potential to detect lymph node metastases (Suzuma et al., 2001). A mathematical model of multifocal tumors suggests that the sum of the tumor sizes across all lesions is the best characteristic which correlates

with the stage and metastatic potential of the tumor (Wodarz et al., 2004). These papers are part of a growing effort to derive a quantitative, mathematical understanding of cancer progression (Moolgavkar and Knudson, 1981; Sheratt and Nowak, 1992; Luebeck and Moolgavkar, 2002; Nowak et al., 2002; Frank et al., 2003; Gatenby and Maini, 2003; Iwasa et al., 2003; Little and Wright, 2003; Michor et al., 2003; Nowak et al., 2003; Iwasa et al., 2004; Michor et al., 2004a,b; Nowak et al., 2004; Michor et al., 2005; Wodarz and Komarova, 2005).

In this paper, we develop a stochastic model to calculate the dynamics of metastasis. We consider cells that acquire metastatic ability by one genetic alteration. This mutation can arise in cancers with different cell numbers (= population sizes). We calculate the expected number of successfully established metastases in dependence of the population size of resident cancer cells, the mutation rates producing mutated cells, and their fitness.

## 2. The model

Consider a population of  $N$  cancer cells proliferating according to the Moran process (Moran, 1962). Initially, all cells are genetically homogeneous—a mutation enabling a cell to metastasize has not yet been produced. At each cell division, a cell is mutated with probability  $u$  to give rise to a cell that has the potential to metastasize. A cell mutated with respect to the metastasis-promoting gene has relative fitness  $r$ . If  $r = 1$ , then the mutation is neutral and the mutated cell has equal proliferation and death rates as resident cancer cells. If  $r > 1$ , the mutation is advantageous and the mutated cell has increased proliferation capabilities or decreased death rates as compared to resident cancer cells. If  $r < 1$ , the mutation is disadvantageous and the mutated cell has decreased proliferation capabilities or increased death rates as compared to resident cancer cells.

At each time step, a cell is chosen for reproduction at random, but proportional to fitness. Here a time unit is equal to the mean time of one cell generation. If there are  $i$  mutated cells, then the probability that a mutated cell is chosen for reproduction is  $ri/(ri + N - i)$ . The chosen cell produces a daughter cell that replaces another randomly chosen cell. The total number of cells remains strictly constant. The probability that a single mutated cell with  $r > 1$  or  $r < 1$  takes over the whole population is given by  $\rho = (1 - 1/r)/(1 - 1/r^N)$  (Moran, 1962). For a neutral mutant,  $r = 1$ , we have  $\rho = 1/N$ . The quantity  $\rho$  is called fixation probability. An advantageous mutation has a higher fixation probability than a neutral mutation, which has a higher fixation probability than a deleterious mutation. The events in a small population, however, are dominated by random drift: if  $N$  is small, then even a deleterious mutation has a certain probability of reaching fixation due to chance events. Here it is assumed that no further mutation occurs while the mutated cell reaches fixation. The transition probabilities of the Moran process

per time interval  $\Delta t$  are given by

$$P(i, i + 1) = \left[ \frac{ri}{1 - i + ri} N(1 - i) \right] \Delta t,$$

$$P(i, i - 1) = \left[ \frac{1 - i}{1 - i + ri} Ni \right] \Delta t,$$

$$P(i, i) = 1 - P(i, i + 1) - P(i, i - 1).$$

Mutated cells are exported at rate  $q$  per cell division to form distant metastatic colonies. Here, we will calculate the rate of metastasis formation and the expected number of successfully established metastases at time  $t$ .

The cumulative risk of producing metastases is given by

$$R = q \int_0^T E(t) dt. \tag{1}$$

Here  $E(t) = N\bar{z}(t)$  denotes the expected number of mutated cells in the tumor, where  $\bar{z}(t)$  is the expected fraction of mutated cells in the cancer at time  $t$ . Let the fraction of resident and mutated cancer cells be  $1 - z$  and  $z$ , respectively. The mean fitness in the cancer is given by  $\bar{w} = 1 - z + rz = 1 + (r - 1)z$ . At each time step, the fraction of mutated cells in the cancer changes due to selection according to

$$(\Delta z)_{sel} = \left( \frac{r}{\bar{w}} - 1 \right) z = \frac{(r - 1)z(1 - z)}{1 + (r - 1)z},$$

and due to mutation according to

$$(\Delta z)_{mut} = u(1 - z).$$

The total rate of change in the fraction  $z$  of mutated cells per time unit, i.e. per cell generation, is given by

$$\frac{dz}{dt} = u(1 - z) + \frac{(r - 1)z(1 - z)}{1 + (r - 1)z} + \sqrt{\frac{2}{N}z(1 - z)} \frac{dB}{dt}. \tag{2}$$

The third term in Eq. (2) indicates the effect of stochastic drift and is represented by a stochastic variable with mean zero and variance  $2z(1 - z)/N$  per time unit. Eq. (2) is a stochastic differential equation interpreted as an Ito-Integral, and the term  $dB/dt$  represents the usual description of white noise.

### 2.1. Exact stochastic computer simulation

We compare our analytical results with direct stochastic computer simulations of the Moran process. At each time step of the Moran process, one cell is chosen for reproduction at random, but proportional to fitness. If there are  $i$  mutated cells with fitness  $r$  in a population of  $N$  cells, then the probability that a mutated cell is chosen for reproduction is  $ri/(ri + N - i)$ . The chosen cell produces a daughter cell, possibly with mutation, that replaces another randomly chosen cell. The total number of cells remains strictly constant. For each parameter choice, we compute many independent runs of the stochastic process. The data points generated by the computer simulation are compared

with the analytical results. Eq. (2), however, does not provide perfect fit with the exact computer simulation of the stochastic process. Therefore, we have to introduce different approximations for the three parameter regimes of the relative fitness of mutated cells.

2.2. Neutral mutation

If the mutation enabling a cell to metastasize is neutral as compared to the resident cancer cells,  $r = 1$ , then the second term of Eq. (2) vanishes. Hence the mean fraction of mutated cells follows:

$$\frac{d\bar{z}}{dt} = u(1 - \bar{z}).$$

The explicit solution of the mean fraction of mutated cells at time  $t$  is given by

$$\bar{z}(t) = 1 - \exp(-ut). \tag{3}$$

The expected number of mutated cells is given by  $E = N\bar{z}(t)$ .

Fig. 2 shows the results of the exact stochastic computer simulation and the numerical simulation of Eq. (3). We plot the expected number of mutated cells,  $E$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size ranges from  $N = 10$  to 100, 1000 and 10 000 in Figs. 2(a)–(d).

2.3. Disadvantageous mutation

If the mutation enabling a cell to metastasize has a fitness disadvantage in the main tumor,  $r < 1$ , then the second term of Eq. (2) is nonlinear and a closed form of  $\bar{z}$  cannot be obtained. The calculation of the average of the selection term requires the variance and higher order moments.

Let  $p(z, t)$  be the probability density of the fraction  $z$  of mutated cells. The diffusion equation corresponding to Eq. (2) has Fokker–Planck form (Arnold, 1973) and is given by

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial z}[f(z)p] + \frac{1}{2}\frac{\partial^2}{\partial z^2}[g(z)p],$$

$$f(z) = \left[ u + \frac{(r-1)z}{1+(r-1)z} \right] (1-z),$$

$$g(z) = \frac{2}{N}z(1-z). \tag{4}$$

Let us now consider the mean and variance of  $z$  using the probability distribution

$$m = \int_0^1 zp(z, t) dz \quad \text{and}$$

$$v = \int_0^1 (z - m)^2 p(z, t) dz. \tag{5}$$

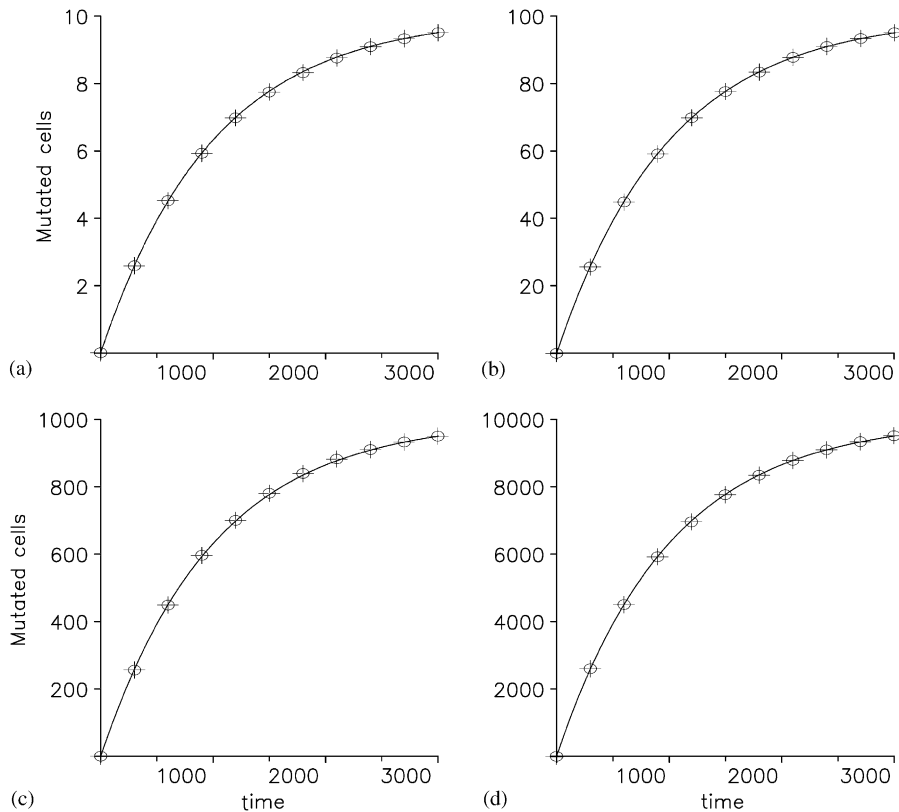


Fig. 2. The mutation conferring metastatic ability is neutral as compared to resident cancer cells,  $r = 1$ . Here we show the results of the exact stochastic computer simulation (circles) and the numerical simulation of Eq. (3). We plot the expected number of mutated cells,  $E = N\bar{z}$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size is  $N = 10$  in (a),  $N = 100$  in (b),  $N = 1000$  in (c) and  $N = 10000$  in (d).

Note that the integrals cover only the interval (0, 1), but the probability for the paths that have already reached fixation is excluded. Denote  $s = 1 - r$ . Then, according to the calculation performed in the Appendix, the moment equations are

$$\frac{dm}{dt} = \left[ u - \frac{sm}{1-sm} \right] (1-m) + \frac{s(1-s)}{(1-sm)^3} v,$$

$$\frac{dv}{dt} = \frac{2}{N} m(1-m) + 2 \left[ -u - \frac{s(1-2m+sm^2)}{(1-sm)^2} - \frac{1}{N} \right] v. \quad (6)$$

The expected fraction of mutated cells in the cancer at time  $t$ ,  $m(t)$ , can be calculated from the initial condition  $m(0) = v(0) = 0$ . The expected number of mutants is given by  $E = Nm(t)$ . This formula, however, does not fit perfectly to the exact computer simulation of the stochastic process. The reason for the deviation is that the derivation of the moment dynamics neglects all trajectories that move out of the open interval (0, 1) (see Appendix). Hence, the probability that a trajectory has already reached fixation,  $z = 1$ , is excluded. For a long time prior to the final fixation, deleterious mutants have a low frequency that is close to the quasi-equilibrium given by the mutation selection balance,  $z = u/(1 - r)$ . Once they reach fixation,

$z = 1$  holds forever. Thus, the probability distribution has two peaks,  $z = m(t)$  and  $z = 1$ , but the moment equation neglects the peak at  $z = 1$ . The value of  $m$  given by Eq. (6) corresponds to the peak at the mutation-selection balance prior to fixation. If we consider the fixation occurring at a random time, the expected number of mutants is given by a mixture of the two peaks of the probability distribution,

$$E = Nm(t)e^{-bt} + N(1 - e^{-bt}). \quad (7)$$

Here  $b = Nu\rho(r)$  and  $\rho(r) = (1 - 1/r)/(1 - 1/r^N)$ . If  $u \ll (1 - r)$ , then we can approximate the expected number of mutants by

$$\bar{z} = m(t) = [u/(1 - r)](1 - \exp[-(1 - r)t]). \quad (8)$$

Fig. 3 shows the results of the exact stochastic computer simulation and the numerical simulation of Eq. (7). We plot the expected number of mutated cells,  $E$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size ranges from  $N = 10$  to 100, 1000 and 10 000 in Figs. 3(a)–(d).

#### 2.4. Advantageous mutation

The moment dynamics given by Eq. (6) is usable only for disadvantageous mutants,  $r < 1$ . If  $r > 1$ , then  $s < 0$  in Eq. (6) and the variance increases exponentially when  $\bar{z}$  is very

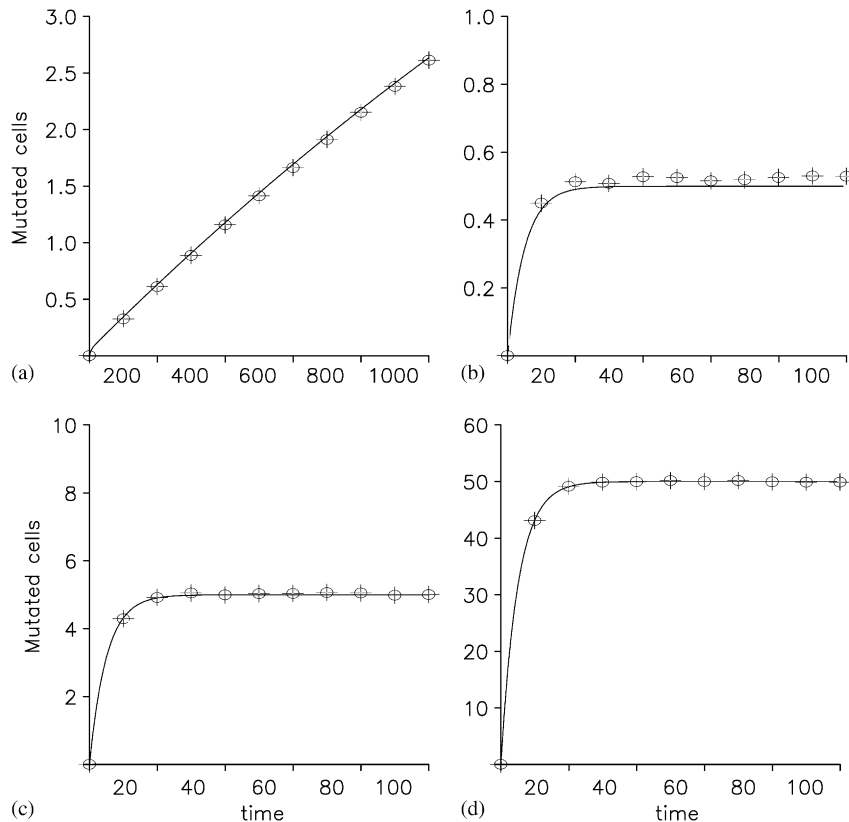


Fig. 3. The mutation conferring metastatic ability is disadvantageous as compared to resident cancer cells,  $r = 0.8$ . We show the results of the exact stochastic computer simulation (circles) and the numerical simulation of Eq. (7). We plot the expected number of mutated cells,  $E$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size is  $N = 10$  in (a),  $N = 100$  in (b),  $N = 1000$  in (c) and  $N = 10000$  in (d).

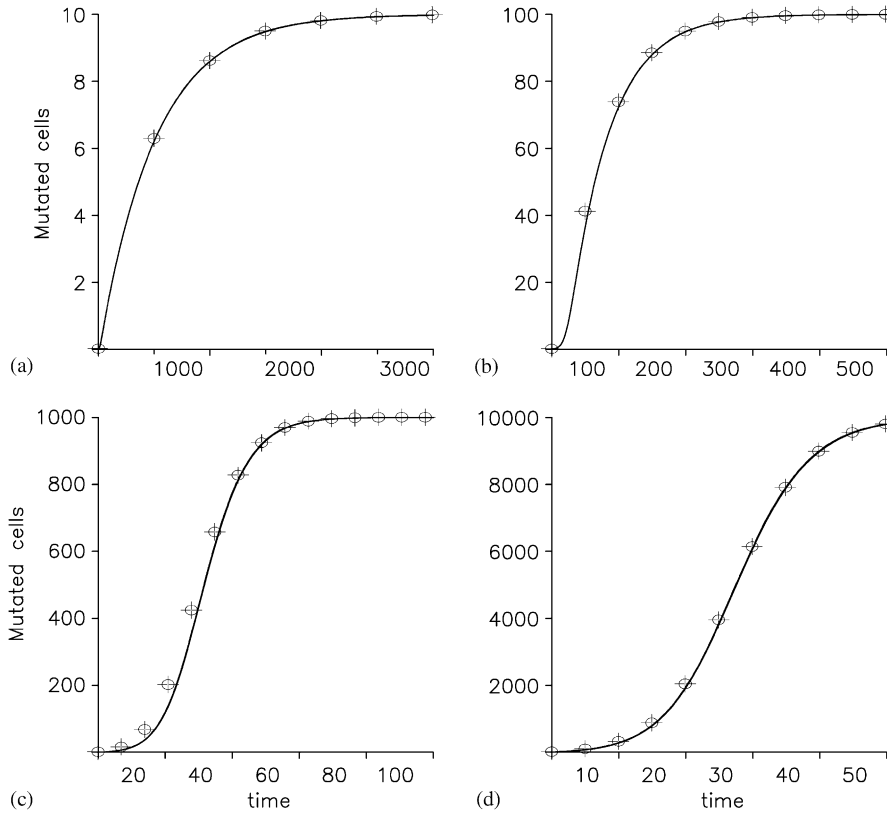


Fig. 4. The mutation conferring metastatic ability is advantageous as compared to resident cancer cells,  $r = 1.2$ . We show the results of the exact stochastic computer simulation (circles) and the numerical simulation of Eq. (9) (line). We plot the expected number of mutated cells,  $E$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size is  $N = 10$  in (a),  $N = 100$  in (b),  $N = 1000$  in (c) and  $N = 10000$  in (d).

small. This leads to negative  $\bar{z}$  when  $r > 1$  but not if  $r < 1$ . Therefore, we consider a different approximation for the dynamics of advantageous mutants in the following.

The evolutionary trajectory of advantageous mutants includes two parts: (i) at low frequency, random drift dominates and stochasticity is important—many lineages of advantageous mutants become extinct, and (ii) at high frequency, selection dominates and the abundance of mutants increases exponentially until reaching fixation. This trajectory is described by the deterministic part of Eq. (2).

The waiting time until the appearance of the first successful mutant follows an exponential distribution with mean  $1/[Nu\rho(r)]$ . The dynamics of the first successful lineage is described by

$$\frac{dz}{dt} = u(1 - z) + \frac{(r - 1)z(1 - z)}{1 + (r - 1)z},$$

where  $z(0) = 1/N$ .

Let  $z(t)$  be the solution of this equation. The expected frequency of mutants in the cancer is given by

$$y(t) = \int_0^t z(t - k)be^{-bk} dk.$$

This frequency is an average over different trajectories in which eventually successful mutants are produced at

different times. Therefore, we have

$$\frac{dy}{dt} = b(z - y). \tag{9}$$

The initial conditions are  $z(0) = 1/N$  and  $y(0) = 0$ . The expected number of mutants is given by  $E = Ny(t)$ .

Fig. 4 shows the results of the exact stochastic computer simulation and the numerical simulation of Eq. (9). We plot the expected number of mutated cells,  $E$ , at time  $t$ . The mutation rate is  $u = 10^{-3}$  per cell division, and the population size ranges from  $N = 10$  to 100, 1000 and 10 000 in Figs. 4(a)–(d).

### 3. Discussion

In this paper, we study the stochastic dynamics of metastasis formation. We consider situations in which metastatic ability is due to a single mutation, such as the activation of an oncogene like MYC or RAS. This mutation can arise in populations of tumor cells of different size and can lead to a fitness advantage, disadvantage or can be neutral as compared to the resident tumor cells. We calculate the expected number of cells with metastatic identity as a function of time, the mutation rate and the fitness value. The number of successfully

established metastases is proportional to the number of mutated cells.

The kinetics of cells with metastatic ability strongly depends on their relative fitness in the main tumor. If cells carrying the metastasis-promoting mutation are disadvantageous when compared to resident cancer cells, then the evolutionary dynamics is described by a jump from a population of resident cancer cells to a population in which mutated cells have reached fixation. Before reaching fixation, mutated cells are maintained at a low level that is determined by the mutation-selection balance. After the fixation, all the cells in the tumor are mutated. If mutated cells are advantageous, then no stable coexistence of resident and mutated cancer cells is possible. Once a mutated cell arises, it is likely to reach fixation in the tumor due to its higher fitness as compared to resident tumor cells. There is, however, a small probability to go extinct even for advantageous mutations. If the mutated cells are perfectly neutral, neither of the above mentioned scenarios is accurate. The mathematical description of all three cases is outlined in the text. Fig. 5 shows the number of mutated cells in dependence of the population size of the main tumor for disadvantageous mutants (Fig. 5a), neutral mutants (Fig. 5b), and advantageous mutants (Fig. 5c). Fig. 6 shows the number of mutated cells in dependence of their fitness.

If the mutation enabling cells to metastasize is advantageous as compared to resident cancer cells, then it is likely to reach fixation in the main tumor such that all cancer cells carry the mutation. If the mutation confers a fitness disadvantage, however, it will be maintained at only a small fraction for a long time (Fig. 7). The number of metastases formed by a tumor also depends on the rate at which mutated cells leave the main tumor and establish metastases elsewhere. This rate is denoted by  $q$ . If all mutations have the same  $q$ , then advantageous mutations are more successful in establishing metastases than neutral mutations, because advantageous mutations are maintained at a higher level in the main tumor. Similarly, neutral mutations are more successful in establishing metastases than disadvantageous mutations. Disadvanta-

geous mutations contribute the same number of metastases as advantageous mutations only if their rate of metastasis formation,  $q$ , is much larger.

Table 1 shows the expected number of metastases generated by a cancer of  $N = 10^6$  cells after one year. For any  $q$ , advantageous mutations produce many more metastases than neutral or disadvantageous mutations. Disadvantageous mutations contribute as many metastases as advantageous mutations only when their  $q$  is a million-fold larger. Therefore, it is very unlikely that a mutation that confers a fitness disadvantage in the main tumor is responsible for metastasis formation.

Our analysis in this paper was motivated by the following question. Is metastatic potential the property of all (or the majority of) cells in the main tumor or only the property of a small subset of cells? We have investigated this question by comparing metastatic mutations that lead

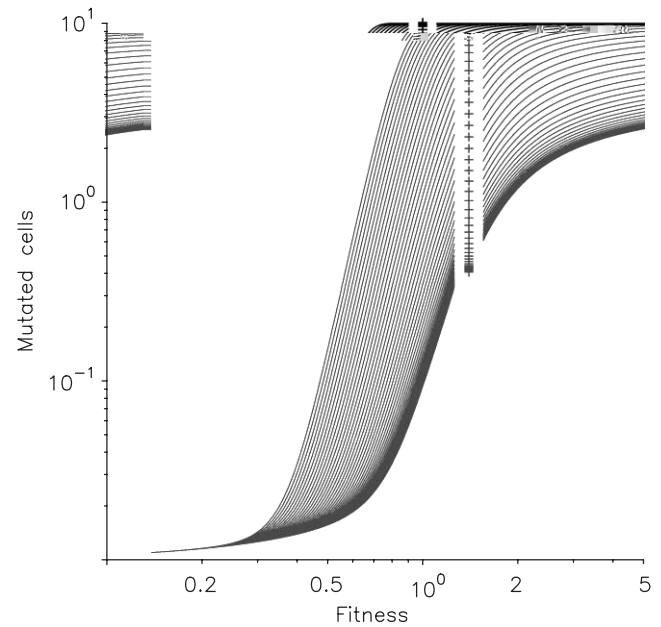


Fig. 6. The figure shows the dependence of the number of mutated cells on their fitness. We show numerical simulation of Eqs. (3) for  $r = 1$ , Eq. (7) for  $r = 0.8$ , and Eq. (9) for  $r = 1.2$ . The lines represent different times.

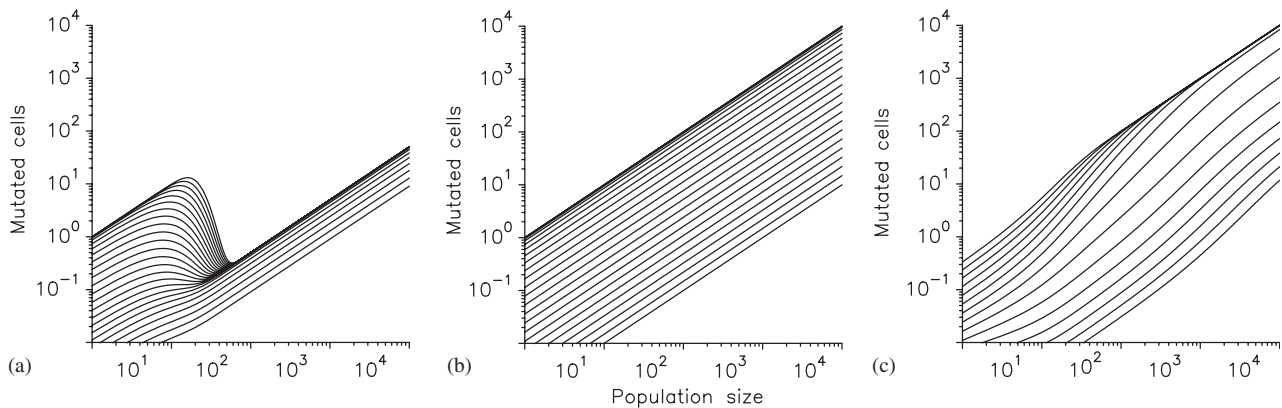


Fig. 5. The figure shows the dependence of the number of mutated cells on the population size of the cancer. (a) shows numerical simulation of Eq. (7) for  $r = 0.8$ , (b) shows Eq. (3) for  $r = 1$  and (c) shows Eq. (9) for  $r = 1.2$ . The lines represent different times, with time flowing upwards in the graph.

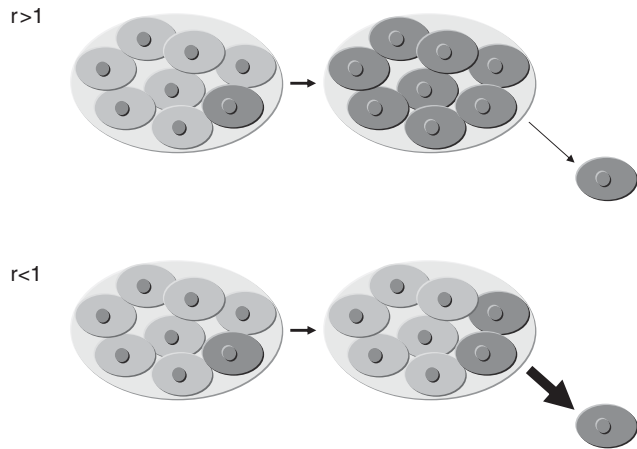


Fig. 7. The number of metastases critically depends on the fitness of a cell carrying a metastasis-promoting mutation in the main tumor. An advantageous mutation,  $r > 1$ , is likely to reach fixation in the main tumor, whereas a disadvantageous mutation,  $r < 1$ , is maintained only at a small fraction. Disadvantageous mutations are successful in establishing metastases only if their rate of export,  $q$ , is very large. Advantageous mutations, in contrast, produce a significant number of metastases even if their rate of export,  $q$ , is small.

Table 1

The expected number of metastases after one year if  $N = 10^6$  for advantageous ( $r = 1.1$ ), neutral ( $r = 1$ ), and disadvantageous ( $r = 0.9$ ) mutations. The mutation rate is  $u = 10^{-8}$  per cell division

|                | $r = 1.1$ | $r = 1$ | $r = 0.9$ |
|----------------|-----------|---------|-----------|
| $q = 1$        | > 1000    | 668     | 36        |
| $q = 0.1$      | > 1000    | 67      | 4         |
| $q = 0.01$     | > 1000    | 7       | 0         |
| $q = 0.001$    | > 1000    | 1       | 0         |
| $q = 0.0001$   | > 1000    | 0       | 0         |
| $q = 0.00001$  | 681       | 0       | 0         |
| $q = 0.000001$ | 68        | 0       | 0         |

Mutated cells are exported at rate  $q$  from the main tumor to form metastases elsewhere. The main tumor has population size  $N = 10^6$ . Advantageous mutations are much more successful at establishing metastases, because they are likely to reach fixation in the main tumor. Neutral and disadvantageous mutations are maintained at low levels in the main tumor. Disadvantageous mutations are as successful in establishing metastases as advantageous mutations only if their rate of export,  $q$ , is a million-fold higher (gray cells). It is therefore very unlikely that mutations that confer a fitness disadvantage in the main tumor are responsible for most metastases produced by the tumor.

either to a selective advantage or to a selective disadvantage in the main tumor. Thus, all our mutations confer some ability to cells to leave the main tumor and induce metastatic growth, but some increase the relative fitness of those cells in the main tumor while others decrease it. The advantageous mutations have a high chance of taking over the entire main tumor. The disadvantageous mutations are more likely to linger at a low frequency (a mutation selection equilibrium) in the main tumor. Therefore, if most metastases derive from advantageous mutations, then it will be the case that all (or the majority of) cells in the

main tumor have metastatic potential. In contrast, if most metastases derive from disadvantageous mutations, then it will be the case that only a small fraction of cells in the main tumor have metastatic potential. Our analyses show that in a large main tumor, advantageous mutations are much more likely to give rise to metastasis production than disadvantageous ones. To compensate for the selective disadvantage in the main tumor, disadvantageous mutations must have a metastatic potential that is many orders of magnitude (a million fold) higher than advantageous mutations. Therefore, according to our calculation most metastases should derive from mutations that have taken over a large part of the main tumor.

There can be exceptions to this rule. If the effective population size of the main tumor is small, then disadvantageous mutations are more effective. It could also be the case that the main tumor is subdivided into spatial compartments, and the metastatic mutation has reached fixation in one spatial compartment but not in the whole main tumor. Finally, it could be the case that there are many more mutations that confer metastatic ability together with a (small) disadvantage in the main tumor, rather than mutations that confer metastatic ability together with the coincidence of being advantageous in the main tumor.

In this paper, we assume that the cancer cell number in the main tumor is approximately constant over time. This assumption describes cancers that are growing slowly or whose cell number is constant until another mutation in an oncogene or tumor suppressor gene has arisen that drives further clonal expansion. It is also possible to assume exponential growth of the cancer. The dynamics of mutations arising in such a scenario are the topic of another paper (Iwasa et al., 2006).

An important goal of the field should be to identify the molecular bases of metastasis and estimate fitness values of cells with metastasis-enabling mutations. Further research into the mechanism of the metastatic process and its individual steps will facilitate quantitative studies of metastasis.

In summary, the dynamics of metastasis formation depend on the mutation rate, population size and fitness value, and different formulas are needed to describe the stochastic dynamics in various parameter regions. Our mathematical framework allows to calculate the expected number of metastases formed by a tumor of constant size, and establishes that metastasis-promoting mutations that confer a fitness disadvantage in the main tumor are unlikely to generate many metastases.

## Acknowledgements

This work was done during Y. I.'s stay at the Program for Evolutionary Dynamics in 2004 and 2005. The Program for Evolutionary Dynamics at Harvard University is sponsored by Jeffrey Epstein.



**Appendix**

*Derivation of the moment dynamics:* Consider the time change in the mean given by Eq. (5), and the Fokker–Planck equation (4). Under the assumption of the zero flux boundary condition (that is, the probability flow is neglected at both ends of the interval), we can integrate in parts as follows:

$$\begin{aligned} \frac{d}{dt} \bar{z} &= \int_0^1 z \frac{\partial}{\partial t} p \, dz = \int_0^1 z \left[ -\frac{\partial}{\partial z} [f(z)p] + \frac{1}{2} \frac{\partial^2}{\partial z^2} [g(z)p] \right] dz \\ &= \int_0^1 f(z)p \, dz - \frac{1}{2} \int_0^1 \frac{\partial}{\partial z} [g(z)p] \, dz \\ &= \int_0^1 f(z)p \, dz. \end{aligned}$$

By using Taylor expansion,  $f(z) = f(\bar{z}) + f'(\bar{z})(z - \bar{z}) + (1/2)f''(\bar{z})(z - \bar{z})^2 + \dots$ , we have

$$\begin{aligned} \frac{d}{dt} \bar{z} &= \int_0^1 \left[ f(\bar{z}) + f'(\bar{z})(z - \bar{z}) + \frac{1}{2} f''(\bar{z})(z - \bar{z})^2 + \dots \right] p \, dz \\ &= f(\bar{z}) + \frac{1}{2} f''(\bar{z})v + \dots \end{aligned}$$

This is equivalent to Eq. (6a). Similarly, we have

$$\begin{aligned} \frac{d}{dt} \langle z^2 \rangle &= \int_0^1 z^2 \frac{\partial}{\partial t} p \, dz \\ &= \int_0^1 z^2 \left[ -\frac{\partial}{\partial z} [f(z)p] + \frac{1}{2} \frac{\partial^2}{\partial z^2} [g(z)p] \right] dz \\ &= \int_0^1 2zf(z)p \, dz - \frac{1}{2} \int_0^1 2z \frac{\partial}{\partial z} [g(z)p] \, dz \\ &= 2 \int_0^1 zf(z)p \, dz + \int_0^1 g(z)p \, dz. \end{aligned}$$

By using the Taylor expansion of  $g(z)$ , we obtain

$$\begin{aligned} \frac{d}{dt} \langle z^2 \rangle &= 2 \int_0^1 z [f(\bar{z}) + f'(\bar{z})(z - \bar{z}) + \frac{1}{2} f''(\bar{z})(z - \bar{z})^2 + \dots] p \, dz \\ &\quad + \int_0^1 [g(\bar{z}) + g'(\bar{z})(z - \bar{z}) + \frac{1}{2} g''(\bar{z})(z - \bar{z})^2 + \dots] p \, dz \\ &= 2[\bar{z}f(\bar{z}) + f'(\bar{z})v + \dots] + g(\bar{z}) + \frac{1}{2} g''(\bar{z})v + \dots \end{aligned}$$

Hence the variance is given by

$$\begin{aligned} \frac{d}{dt} v &= \frac{d}{dt} \langle z^2 \rangle - 2\bar{z} \frac{d}{dt} \bar{z} \\ &= 2 \left[ \bar{z}f(\bar{z}) + f'(\bar{z})v + \frac{1}{2} f''(\bar{z})\langle z(z - \bar{z})^2 \rangle + \dots \right] \\ &\quad + g(\bar{z}) + \frac{1}{2} g''(\bar{z})v + \dots - 2\bar{z} \left[ f(\bar{z}) + \frac{1}{2} f''(\bar{z})v \right] \\ &= 2f'(\bar{z})v + g(\bar{z}) + \frac{1}{2} g''(\bar{z})v + \dots \end{aligned}$$

Here third order and other moments are neglected in order to close the dynamics. The last equation gives Eq. (6b).

With Eq. (4) and  $s = 1 - r > 0$ , we have

$$f'(z) = -u - \frac{s(1 - 2z + sz^2)}{(1 - sz)^2},$$

$$f''(z) = \frac{2s(1 - s)}{(1 - sz)^3},$$

$$g''(z) = -4/N,$$

which leads to Eq. (6) in the text.

**References**

Arnold, L., 1973. Stochastic Differential Equation: Theory and Applications. Wiley, New York.

Bernards, R., Weinberg, R.A., 2002. A progression puzzle. *Nature* 418, 823.

Bosl, G.J., Geller, N.L., Cirrincione, C., Vogelzang, N.J., Kennedy, B.J., Whitmore, W.F., Vugrin, D., Scher, H., Nisselbaum, J., Golbey, R.B., 1983. Multivariate analysis of prognostic variables in patients with metastatic testicular cancer. *Cancer Res.* 43, 3403–3407.

Chambers, A., Naumov, G., Vantuyghem, S., Tuck, A., 2000. Molecular biology of breast cancer biology: clinical implications of experimental studies on metastatic inefficiency. *Breast Can. Res.* 2, 400–407.

Chambers, A., Naumov, G.N., Varghese, H.J., Nadkarni, K.V., MacDonald, I.C., Groom, A.C., 2001. Critical steps in hematogenous metastasis: an overview. *Surg. Oncol. Clin. North Am.* 10, 243–255.

Chambers, A., Groom, A., MacDonald, I., 2002. Dissemination and growth of cancer cells in metastatic sites. *Nat. Rev. Cancer* 2, 563–572.

Delsanto, P.P., Romano, A., Scalerandi, M., Pescarmona, G.P., 2000. Analysis of a “phase transition” from tumor growth to latency. *Phys. Rev. E* 62, 2547–2554.

Fidler, I., 2002. Critical determinants of metastasis. *Semin. Cancer Biol.* 12, 89–96.

Frank, S.A., Iwasa, Y., Nowak, M.A., 2003. Patterns of cell division and the risk of cancer. *Genetics* 163, 1527–1532.

Gatenby, R.A., Maini, P.K., 2003. Mathematical oncology: cancer summed up. *Nature* 421, 321.

Iwasa, Y., Michor, F., Nowak, M.A., 2003. Evolutionary dynamics of escape from biomedical intervention. *Proc. R. Soc. London B* 270, 2573–2578.

Iwasa, Y., Michor, F., Nowak, M.A., 2004. Stochastic tunnels in evolutionary dynamics. *Genetics* 166, 1571–1579.

Iwasa, Y., Nowak, M.A., Michor, F., 2006. Evolution of resistance during clonal expansion. *Genetics*, in press.

Little, M.P., Wright, E.G., 2003. A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Math. Biosci.* 183, 111–134.

Luebeck, E.G., Moolgavkar, S.H., 2002. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc. Natl Acad. Sci. USA* 99, 15095–15100.

Michor, F., Iwasa, Y., Komarova, N.L., Nowak, M.A., 2003. Local regulation of homeostasis favors chromosomal instability. *Curr. Biol.* 13, 581–584.

Michor, F., Iwasa, Y., Nowak, M.A., 2004a. Dynamics of cancer progression. *Nat. Rev. Can.* 4, 197–205.

Michor, F., Iwasa, Y., Rajagopalan, H., Lengauer, C., Nowak, M.A., 2004b. Linear model of colon cancer initiation. *Cell Cycle* 3, 358–362.

Michor, F., Iwasa, Y., Hughes, T.P., Branford, S., Shah, N.P., Sawyers, C.L., Nowak, M.A., 2005. Dynamics of chronic myeloid leukemia. *Nature* 435, 1267–1270.

Moolgavkar, S.H., Knudson, A.G., 1981. Mutation and cancer: a model for human carcinogenesis. *J. Natl Cancer Inst.* 66, 1037–1052.

Moran, P., 1962. The Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

- Nowak, M.A., Komarova, N.L., Sengupta, A., Jallepalli, P.V., Shih, I.M., Vogelstein, B., Lengauer, C., 2002. The role of chromosomal instability in tumor initiation. *Proc. Natl Acad. Sci. USA* 99, 16226–16231.
- Nowak, M.A., Michor, F., Iwasa, Y., 2003. The linear process of somatic evolution. *Proc. Natl Acad. Sci. USA* 100, 14966–14969.
- Nowak, M.A., Michor, F., Komarova, N.L., Iwasa, Y., 2004. Evolutionary dynamics of tumor suppressor gene inactivation. *Proc. Natl Acad. Sci. USA* 101, 10635–10638.
- Panetta, J.C., 1996. A mathematical model of periodically pulsed chemotherapy: tumor recurrence and metastasis in a competitive environment. *Bull. Math. Biol.* 58, 425–447.
- Pescarmona, G.P., Scalerandi, M., Delsanto, P.P., Condat, C.A., 1999. Nonlinear model of cancer growth and metastasis: a limiting nutrient as a major determinant of tumor shape and diffusion. *Med. Hypotheses* 53, 497–503.
- Pozzatti, R., Muschel, R., Willians, J., et al., 1986. Primary rat embryo cells transformed by one or two oncogenes show different metastatic potentials. *Science* 232, 223–227.
- Ramaswamy, S., Ross, K.N., Lander, E.S., Golub, T.R., 2003. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* 33, 49–54.
- Sheratt, J.A., Nowak, M.A., 1992. Oncogenes, anti-oncogenes and the immune response to cancer: a mathematical model. *Proc. R. Soc. London B* 248, 261–271.
- Sorlie, T., Perou, C.M., Tibshirani, R., et al., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* 98, 10869–10874.
- Steege, P.S., Bevilacqua, G., Kopper, L., Thorgerisson, U.P., Talmadge, J.E., Liotta, L.A., Sobel, M.E., 1988. Evidence for a novel gene associated with low tumor metastatic potential. *J. Natl Cancer Inst.* 80, 200–204.
- Suzuma, T., Sakurai, T., Yoshimura, G., Umemura, T., Tamaki, T., Naito, Y., 2001. A mathematical model of axillary lymph node involvement considering lymph node size in patients with breast cancer. *Breast Cancer* 8, 206–212.
- Thames, H.D., Buchholz, T.A., Smith, C.D., 1999. Frequency of first metastatic events in breast cancer: implications for sequencing of systemic and local-regional treatment. *J. Clin. Oncol.* 17, 2649–2658.
- van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H., Bernards, R., 2002. A gene expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999–2009.
- Veer, L.J.V., Weigelt, B., 2003. Road map to metastasis. *Nat. Med.* 9, 999–1000.
- Veer, L.J.V., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H., 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.
- Weiss, L., 2000. Metastasis of cancer: a conceptual history from antiquity to the 1990s. *Cancer Metastasis Rev.* 19, 193–383.
- Wodarz, D., Komarova, N., 2005. *Computational Biology of Cancer: Lecture Notes and Mathematical Modeling*. World Scientific, Singapore.
- Wodarz, D., Krakauer, D.C., 2001. Genetic instability and the evolution of angiogenic tumor cell lines. *Oncol. Rep.* 8, 1195–1201.
- Wodarz, D., Iwasa, Y., Komarova, N.L., 2004. On the emergence of multifocal cancers. *J. Carcinogen* 3, 13.
- Wyllie, A.H., Rose, K.A., Morris, R.G., Steel, C.M., Foster, E., Spandidos, D.A., 1987. Rodent fibroblast tumors expressing human myc and ras genes: growth, metastasis and endogenous oncogene expression. *Br. J. Cancer* 56 (3), 251–259.
- Yu, D.H., Hung, M.C., 1991. Expression of activated rat neu oncogene is sufficient to induce experimental metastasis in 3t3 cells. *Oncogene* 6, 1991–1996.