OPEN ACCESS
IOP Publishing
Physical Biology
Phys. Biol. 11 (2014) 046003 (13pp)
doi:10.1088/1478-3975/11/4/046003

Quantifying metastatic inefficiency: rare genotypes versus rare dynamics

Luis H Cisneros[1] **and Timothy J Newman**[1,2]

[1] Center for Convergence of Physical Science and Cancer Biology, Arizona State University, Tempe, AZ 85287, USA
[2] Division of Cell and Developmental Biology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK

E-mail: T.Newman@dundee.ac.uk

Received 16 March 2014
Accepted for publication 18 June 2014
Published 17 July 2014

Abstract

We introduce and solve a 'null model' of stochastic metastatic colonization. The model is described by a single parameter θ: the ratio of the rate of cell division to the rate of cell death for a disseminated tumour cell in a given secondary tissue environment. We are primarily interested in the case in which colonizing cells are poorly adapted for proliferation in the local tissue environment, so that cell death is more likely than cell division, i.e. $\theta < 1$. We quantify the rare event statistics for the successful establishment of a metastatic colony of size N. For $N \gg 1$, we find that the probability of establishment is exponentially rare, as expected, and yet the mean time for such rare events is of the form $\sim \log(N)/(1-\theta)$ while the standard deviation of colonization times is $\sim 1/(1-\theta)$. Thus, counter to naive expectation, for $\theta < 1$, the average time for establishment of successful metastatic colonies decreases with decreasing cell fitness, and colonies seeded from lower fitness cells show less stochastic variation in their growth. These results indicate that metastatic growth from poorly adapted cells is rare, exponentially explosive and essentially deterministic. These statements are brought into sharper focus by the finding that the temporal statistics of the early stages of metastatic colonization from low-fitness cells ($\theta < 1$) are statistically indistinguishable from those initiated from high-fitness cells ($\theta > 1$), i.e. the statistics show a duality mapping $(1-\theta) \to (\theta - 1)$. We conclude our analysis with a study of heterogeneity in the fitness of colonising cells, and describe a phase diagram delineating parameter regions in which metastatic colonization is dominated either by low or high fitness cells, showing that both are plausible given our current knowledge of physiological conditions in human cancer.

Keywords: cancer, metastasis, rare events, stochastic processes

(Some figures may appear in colour only in the online journal)

1. Introduction

Cancer metastasis is the process of dissemination of cancer cells originating from a solid tumour, and their subsequent colonization of distant tissues [1, 2]. Metastatic disease is responsible for the great majority of cancer deaths. It is generally considered to be a succession of unlikely events rendering *in toto* a highly inefficient process [3–5]. In the conventional view of metastasis, only certain cancer cells from the primary tumour can successfully metastasise as they are thought to require a number of attributes, such as the ability to invade local tissue, enter, survive in, and leave the bloodstream, and colonize new tissue environments. As such, they have been likened to decathletes, in that they possess multiple pre-adapted abilities, all of which are required to form a new colony in a distant tissue [6]. This view explains metastatic inefficiency, as it highlights the low probability of

acquiring all the specific genetic mutations necessary to achieve the intermediate goals *en route* to successful metastasis [7, 8]. However, as pointed out by Bernards and Weinberg [9], this picture contains a conceptual difficulty concerning the microevolution of such cells. The acquisition of these abilities would appear to endow the cells with no selective advantage in the primary tumour, and, as such, one might expect the sub-population of fully-fledged metastatic cells to be vanishingly small. The resolution put forward by Bernards and Weinberg posits pleiotropy, namely, that genetic mutations enabling metastasis might also provide selective advantage in the primary tumour, but this view is then hard to reconcile with metastatic inefficiency. An additional and important piece of information is the observation of large numbers of circulating tumour cells (CTCs) and disseminated tumour cells (DTCs) in cancer patients with no discernible metastatic disease [10–12], which provides a challenging backdrop in which to understand metastatic inefficiency. Clearly, there is scope for a more precise and quantitatively based understanding of metastasis, the first steps of which we hope to provide in this paper.

Given the large number of CTCs in the blood stream and DTCs in the bone marrow and other secondary sites of cancer patients, one can posit a large number of individual attempts at metastatic colonization occurring over time, most of which are unsuccessful; so that the inefficiency of metastasis can be approached as a problem in the statistics of rare events. For the purposes of this paper, we delineate such rare events into two types. First, rare genotypes, as considered in the mainstream paradigm; it being assumed that there exists only a minute subset of colonising cancer cells capable of successful metastasis. Second, rare dynamics, which is the concept we wish to explore in depth in this paper; in which we posit that highly unlikely colonization dynamics can arise from the statistical fluctuations of birth and death of otherwise poorly adapted seeding cells. A helpful metaphor to understand the difference between these two types of rare events is the use of special forces versus infantry to achieve a military mission. 'Rare genotypes' corresponds to a highly trained yet small group of soldiers who each have the specialized skills to accomplish the mission. 'Rare dynamics' corresponds to a large number of poorly trained infantry, none of whom have the specialized skills required, but each of whom has a very small but non-zero chance of accomplishing the mission. We will study these rare events in the context of the final stage of metastasis, that of tissue colonization, which is arguably the most poorly understood step of metastasis, and yet observed to be the highest barrier for DTCs to overcome in forming a new tumour [2, 4, 5]; in mouse models it is observed that the vast majority of metastatic cells readily die once they reach a distant tissue site and attempt to colonize it [13]. We propose a simple birth-death model of stochastic colonization which will allow us to quantify the temporal statistics of rare dynamics. The results arising from this model are counter-intuitive at first glance, and provide a fascinating duality between colonization arising from rare genotypes and rare dynamics. We then proceed to quantify the relative likelihood of colonization from these two different sources, based on human cancer data to hand, and show plausibility that metastatic colonization can equally well arise purely by chance, not requiring mutations of key genes.

2. Methods

We are interested in the probability and dynamics of metastatic colonization, namely, the process by which a DTC, having successfully become located in a secondary tissue environment, begins proliferation to attempt to create a new tumour (a micro-metastasis). We will approach this problem as a stochastic rather than a deterministic process, namely, we ascribe to the founding cell a rate (probability per unit time) of cell death μ, and, similarly a rate of cell division r. The key parameter in the model is the ratio of these two rates, which we denote by $\theta = r/\mu$, and which we term 'fitness'. It is very important to understand that the fitness is a property of a given cell in a given environment (i.e. not a property that can be ascribed to a cell *per se*). Later in our analysis, we will consider heterogeneity of fitness in the DTC population in order to discuss how our model relates to Paget's classical seed and soil hypothesis [6], which is used to explain to correlation of primary and secondary sites across different cancer types. Given that we are interested in the early stages of secondary tumour growth, we can assume that the birth and death rates are independent of the number of cells in the colony (i.e. the growth and death rates are density-independent). Thus, the stochastic process defining our model is a linear birth/death process [14]. We assume that the realizations of this process, i.e. the seeding events occurring in secondary sites by DTCs, are statistically independent. We also assume that progeny cells have the same fitness as the founding cell. This assumption will clearly break down, with interesting consequences, for a long-lived colony of cells, but will be a good approximation for the nascent micro-tumours of interest here. We note here that there is prior work on applying stochastic processes to tumour growth, but these have all been concerned with growth arising from high fitness cells in a microevolutionary framework [15–17].

The fitness parameter θ encapsulates all the factors affecting the proliferative potential of single cells in a particular microenvironment (vulnerability to apoptosis, interaction with stroma and immune system, etc). For cases when $\theta > 1$, a founding DTC is more likely to divide than die, as are its progeny, and so it is likely that a tumour will arise, following an approximately exponential growth trajectory. Conversely, if $\theta < 1$ cells tend to die more often than dividing, and the overwhelmingly most likely outcome of each seeding is that the nascent secondary tumour will rapidly succumb to extinction. However, rare events can occur in which a nascent colony can achieve a critical size, which, as we describe shortly, we define to be quasi-stable. A primary concern in this paper is to calculate the absolute probability and associated temporal statistics of these rare dynamical events.

If cells remain poorly adapted to their environment, with $\theta < 1$, every attempt to form a sustainable tumour will ultimately fail given long enough time, since zero cells is the only

absorbing state for this stochastic process, as described. Biologically, there are at least two ways in which a fragile tumour arising from a low-fitness founding cell can evade extinction. First, by reaching a critical size N that affords a less hostile internal microenvironment to the cells. In a tumour of this moderate size, cells internal to the tumour, which are not in direct contact with the hostile microenvironment, will have a higher effective fitness, $\theta' > 1$, rendering the tumour quasi-stable and allowing essentially deterministic growth thereafter. Second, given that cancer cells are genetically unstable, each division event carries a chance to produce daughter cells of a slightly different genotype, which after several generations provides the possibility, through microevolution, of adapted cells arising in the tissue niche. In either case, there is a critical threshold that new colonies have to surpass in order to become stable. The first (microenvironment) case requires a critical colony size N to be reached, while the second (microevolution) case requires a critical number of division events to allow significant diversity to arise and thus the possibility for adaptation. We shall address here the first case, assuming cells to retain the fitness of the founding cells, which already provides rich and interesting results. The statistical properties of the second case are more complicated and are not within the purview of this paper. In brief, the basic assumptions of our birth/death model are:

- All colonization attempts start with single DTCs and are independent of each another.
- All cells within a nascent tumour have the same fitness. Dynamical adaptation and microevolution are not considered.
- Cell divisions and deaths occur stochastically according to Poisson statistics.
- Only small nascent metastases, not limited by nutrients, are considered. As such, the rates of cell division and death are assumed independent of the cluster size.
- For metastases arising from a low-fitness cell, there is a critical size N at which nascent tumours render a new internal microenvironment, shielding inside cells from the deleterious effects of the external tissue, and stabilising the stochastic growth.

These assumptions are sufficient for us to formulate this model as a stochastic process which can be analysed using a linear master equation, as described below in some detail. At first glance one would imagine that a simple model such as this would have been studied in detail decades ago. To our knowledge this is not the case, although we note that Basan *et al* [18] studied a non-linear version of this model in a biomechanical study of metastatic growth, wherein forces within the host tissue act to oppose tumour growth until a critical mass is achieved. Similar models of stochastic colonization have been studied in island ecology, with the biological assumption that new individuals are pre-adapted to the new habitat ($\theta > 1$), and thus most likely proliferate [19]. The interest of such studies resides in calculating the statistics of small population size extinctions wiping out strong, but fledgling colonies. This reasoning has also been applied to stochastic re-colonizations of pre-adapted metastatic cancer stem cells during treatment cycles [17].

We use well-known analytical methods of stochastic process theory [14] to construct and solve our model. Solutions for temporal statistics of interest can be calculated exactly, albeit after lengthy calculations, using generating functions and Laplace transform techniques. Furthermore, our results are fully supported by numerical calculations and simulations. These simulations provide us with explicit rare event realizations that are helpful in guiding our understanding of some of the rather counterintuitive results that we report.

First, we consider the colonization process arising from a population of DTCs with the same fitness θ. These results will be used later when considering heterogeneity. We define the function $P(n, t)$ to be the probability that the nascent tumour has n cells at time t, given that there existed a single cell at time $t = 0$. This probability increases when a division event occurs in the $(n-1)$-state or a cell death event occurs in the $(n+1)$-state. Likewise, $P(n, t)$ will decrease when division or death events occur in the n-state. The rate of change of this probability function is given by the master equation:

$$\frac{\mathrm{d}P(n,t)}{\mathrm{d}t} = (n+1)P(n+1,t) - nP(n,t) + \theta\left[(n-1)P(n-1,t) - nP(n,t)\right], \quad (1)$$

defined for $n \geqslant 0$ with the understanding that

$$\frac{\mathrm{d}P(0,t)}{\mathrm{d}t} = P(1,t), \quad (2)$$

and the single-seeding initial condition $P(n, 0) = \delta_{n,1}$. We have scaled time by the cell death rate μ for convenience.

In the terminology of stochastic process theory, $n = 0$ is an absorbing state. Since our primary interest is in low-fitness cells ($\theta < 1$), and we assume the existence of a critical size N, we will study the statistics of first passage by imposing a second absorbing state at $n = N$ [20]. This is accomplished by limiting equation (1) to $0 \leqslant n \leqslant (N-2)$, and disallowing death events from the state $n = N$:

$$\frac{\mathrm{d}P(N-1,t)}{\mathrm{d}t} = -(N-1)P(N-1,t) + \theta\left[(N-2)P(N-2,t) - (N-1)P(N-1,t)\right], \quad (3)$$

$$\frac{\mathrm{d}P(N,t)}{\mathrm{d}t} = \theta(N-1)P(N-1,t). \quad (4)$$

It is helpful to introduce $Q_N(t)\delta t$ as the probability of reaching the critical size N in the time interval $(t, t+\delta t)$, from which we have

$$Q_N(t) = \frac{\mathrm{d}P(N,t)}{\mathrm{d}t} = \theta(N-1)P(N-1,t). \quad (5)$$

For fitness values $\theta > 1$ the existence of N is of little biological relevance, but one can still ask the corresponding statistical question: what are the first-passage temporal statistics for colonies arising from pre-adapted cells reaching size N? As such, we will use the same model to discuss nascent tumour growth for colonies arising from both low and high

fitness founding DTCs, since the details of the calculations are insensitive to the size of θ.

Solving the $(N+1)$ differential equations (1)–(4) yields $P_N(t)$ and $Q_N(t)$, from which all statistics of successful colonization can be derived. In particular we are interested in the following three quantities:

(i) The probability of successful establishment of a colony of size N:

$$P_N = \int_0^\infty \mathrm{d}t\ Q_N(t), \tag{6}$$

(ii) The average time to reach size N in the subset of successful realizations:

$$\langle t_N \rangle = \frac{\int_0^\infty \mathrm{d}t\ t\ Q_N(t)}{\int_0^\infty \mathrm{d}t\ Q_N(t)}, \tag{7}$$

(iii) The variance in the time to reach size N in the subset of successful realizations:

$$\mathrm{var}(t_N) = \left\langle t_N^2 \right\rangle - \left\langle t_N \right\rangle^2, \tag{8}$$

with

$$\langle t_N^2 \rangle = \frac{\int_0^\infty \mathrm{d}t\ t^2\ Q_N(t)}{\int_0^\infty \mathrm{d}t\ Q_N(t)}. \tag{9}$$

The Laplace transform method is used to solve equations (1)–(4) in order to obtain $Q_N(t)$, and then calculate these statistics. Because our primary interest is in non-pre-adapted cells ($\theta < 1$) we shall present some results in a form that it is most appropriate for this case, but the derivation and general results are valid for arbitrary θ values.

3. Results and discussion

3.1. Probability of establishment

We provide details of the calculation of various statistical quantities, starting with the probability of establishment of a micrometastasis of size N. By defining $\hat{P}_n(s)$ to be the Laplace transform of $P_n(t)$, the master equations and boundary conditions take the form of the following $N+1$ algebraic equations:

$$s\hat{P}_0 = \hat{P}_1, \tag{10}$$

$$s\hat{P}_1 = 2\hat{P}_2 - (1+\theta)\hat{P}_1 + 1, \tag{11}$$

$$s\hat{P}_2 = 3\hat{P}_3 - 2(1+\theta)\hat{P}_2 + \theta\hat{P}_1, \tag{12}$$

$$\vdots$$

$$s\hat{P}_n = (n+1)\hat{P}_{n+1} - n(1+\theta)\hat{P}_n + (n-1)\theta\hat{P}_{n-1}, \tag{13}$$

$$\vdots$$

$$s\hat{P}_{N-1} = -(N-1)(1+\theta)\hat{P}_{N-1} + (N-2)\theta\hat{P}_{N-2}, \tag{14}$$

$$s\hat{P}_N = (N-1)\theta\hat{P}_{N-1}. \tag{15}$$

Equations (11)–(14) form a closed set of $N-1$ equations and can be solved explicitly.

We now define the set of functions $\{H_n(s)\}$ via the recursion relation

$$H_n(s) = \frac{\theta}{s/n + \theta + 1 - H_{n-1}(s)}, \tag{16}$$

with $H_0 = 0$. Then solving the algebraic equations (11)–(13) iteratively, one finds

$$\theta n\hat{P}_n = H_1 \cdots H_n + (n+1)\hat{P}_{n+1}H_n. \tag{17}$$

Setting $n = N-2$ in (17) and eliminating $\hat{P}_{N-2}$ using (14) and then (5) we obtain a simple expression for $\hat{Q}_N(s)$:

$$\hat{Q}_N(s) = H_1 \cdots H_{N-1}. \tag{18}$$

It is subsequently convenient here to work with the set of functions $\{J_n(s)\}$ defined by

$$J_n(s) = \theta^n / \left(H_1 \cdots H_n\right). \tag{19}$$

Then $H_n = \theta J_{n-1}/J_n$ and using (16) we get the recursion relation:

$$J_n(s) = \left(\frac{s}{n} + 1 + \theta\right) J_{n-1}(s) - \theta J_{n-2}(s), \tag{20}$$

with $J_0 \equiv 1$ and $J_1 = s + 1 + \theta$. The function of interest is therefore

$$\hat{Q}_N(s) = \frac{\theta^{N-1}}{J_{N-1}(s)}. \tag{21}$$

Now, using the definition of the Laplace transform, the integrals of $Q_N(t)$ with weights 1, t, and t^2 are respectively given by $\hat{Q}_N(0)$, $-\hat{Q}_N'(0)$, and $\hat{Q}_N''(0)$. Therefore the probability of successful establishment (equation (6)) is

$$P_N = \int_0^\infty \mathrm{d}t\ Q_N(t) = \hat{Q}_N(0) = \frac{\theta^{N-1}}{J_{N-1}(0)}. \tag{22}$$

From equation (20) we have

$$J_n(0) = (1+\theta)J_{n-1}(0) - \theta J_{n-2}(0), \tag{23}$$

with $J_0(0) = 1$ and $J_1(0) = 1 + \theta$. This recursion relation may be iterated to find the simple result

$$J_n(0) = 1 + \theta + \theta^2 + \ldots \theta^n \tag{24}$$

thus

$$J_{N-1}(0) = 1 + \theta + \theta^2 + \ldots \theta^{N-1} = \frac{\left(1-\theta^N\right)}{(1-\theta)}. \tag{25}$$

Substituting in (22) we have our first main result, viz. the

probability of establishment is given by:

$$P_N = \frac{(1-\theta)\theta^{N-1}}{\left(1-\theta^N\right)}. \tag{26}$$

3.2. Mean and variance of colonization times

The temporal statistics can be obtained in a similar manner. From (21) we have that

$$\langle t_N \rangle = \frac{\int_0^\infty \mathrm{d}t\ t\ Q_N(t)}{\int_0^\infty \mathrm{d}t\ Q_N(t)} = -\frac{\hat{Q}_N'(0)}{\hat{Q}_N(0)} = \frac{J_{N-1}'(0)}{J_{N-1}(0)}. \tag{27}$$

Differentiating (20) with respect to s and setting $s = 0$ one finds

$$J_n'(0) = (1+\theta)J_{n-1}'(0) - \theta J_{n-2}'(0) + J_n(0)/n. \tag{28}$$

Solving by iteration for $J_n'(0)$ and using equation (27) gives an exact solution for the average time of colonization:

$$\langle t_N \rangle = \frac{1}{(1-\theta)} \sum_{m=1}^{N} \frac{\left(1-\theta^m\right)\left(1-\theta^{N-m}\right)}{m\left(1-\theta^N\right)}. \tag{29}$$

Following the same procedure, in solving by iteration for $J_n(0)''$, one finds after a lengthy calculation an exact solution for the variance in the time of colonization:

$$\mathrm{var}(t_N) = \langle t_N \rangle^2 - \frac{2}{(1-\theta)^2} \sum_{m=2}^{N-1} \frac{\left(1-\theta^{N-m}\right)}{m\left(1-\theta^N\right)} \times \sum_{m'=1}^{m-1} \frac{\left(1-\theta^{m'}\right)\left(1-\theta^{m-m'}\right)}{m'}. \tag{30}$$

3.3. Asymptotic results for large N

We are interested in micrometastases of moderate size, so it is useful to extract the asymptotic form of the temporal statistics for $N \gg 1$. Equation (26) gives the probability of establishment of a tumour of size N for any fitness θ. For $\theta < 1$, but not too close to unity (i.e. $1-\theta \gg 1/N$), this equation takes the simple exponential form:

$$P_N \sim a\theta^N, \tag{31}$$

with prefactor $a = (1-\theta)/\theta$. Quite intuitively, the probability of establishment decreases exponentially for increasing N. For example, a micro-tumour of 20 cells with fitness of $\theta = 0.5$ has a probability of forming $P_{20} \sim (0.5)^{20} \sim 10^{-6}$. Hence, one in a million progenitor cells will achieve this size for this fitness level.

The expressions for the average and variance in colonization times are too complicated to easily infer the dependence on the fitness θ and critical size N. By recasting the summations as integrals and performing an asymptotic analysis for $N \gg 1$ it is possible to extract the leading behaviour, which turns out to be quite simple. In these calculations, it is convenient to use the integral representation

$$\frac{1}{m} = \int_0^\infty \mathrm{d}u\ e^{-mu} \tag{32}$$

in order to explicitly perform the summations, which have the form of a geometric series.

For the average colonization time we find, for $\theta < 1$ and not too close to unity,

$$\langle t_N \rangle = \frac{\log N}{(1-\theta)} + \frac{(\gamma + \log(1-\theta))}{(1-\theta)} + O(1/N), \tag{33}$$

where $\gamma = 0.57721\ldots$ is Euler's constant [21]. Because of the logarithmic dependence of the average colonization time on N, we can rewrite this expression as a simple exponential growth law:

$$N = A(\theta) \exp\left(g_{\mathrm{eff}} \langle t_N \rangle\right), \tag{34}$$

where $A(\theta) = e^{-\gamma}/(1-\theta)$ and the effective growth rate is $g_{\mathrm{eff}} = (1-\theta)$. Thus, we find that the average growth dynamics is *exponential* with a rate *anti-correlated* to the fitness. In other words, the less fit the cells are, the more aggressive the colonization dynamics of (rare) successful tumours will appear to be. On first reading, this result is counter-intuitive. One might imagine that the rare dynamics leading to colonies of low-fitness cells would be slow growing, relying on repeated chance avoidance of extinction. Despite the huge combinatorial weight of such meandering trajectories, the statistically relevant successful trajectories are those which grow rapidly, relying on repeated birth events and a relative absence of death events. This extreme bias becomes ever more important as θ decreases, hence the more rapid successful trajectories for cells of lower fitness.

Similar counter-intuitive results of this type are known in disparate fields, such as the 'bold play' strategy in gambling (successive 'all or nothing' gambling on high odds games is an optimal strategy for rare but large wins) [22], and transition rate theory in chemistry to quantify spontaneous dissociation of chemical bonds (stronger bonds break rarely but explosively). The Freidlin–Wentzell theory of large deviations provides a general mathematical framework to further understand such rare event dynamics, including the time-reversal duality discussed below [23]. We are obliged to ask whether this average time is a meaningful statistical measure. Perhaps the fluctuations in successful realizations are so great that the average time is a statistical oddity, with little bearing on the true nature of the dynamics. The variance of the distribution is informative in this regard.

Performing an asymptotic analysis of equation (30), and after a lengthy calculation, we find that the variance in colonization times is given by

$$\mathrm{var}(t_N) = \frac{\zeta(2)}{(1-\theta)^2} + O(1/N), \tag{35}$$

where the Riemann Zeta function $\zeta(2) = \pi^2/6$ [21]. Here we see that for large N the variance is independent of the critical tumour size and decreases with θ. This means that successful

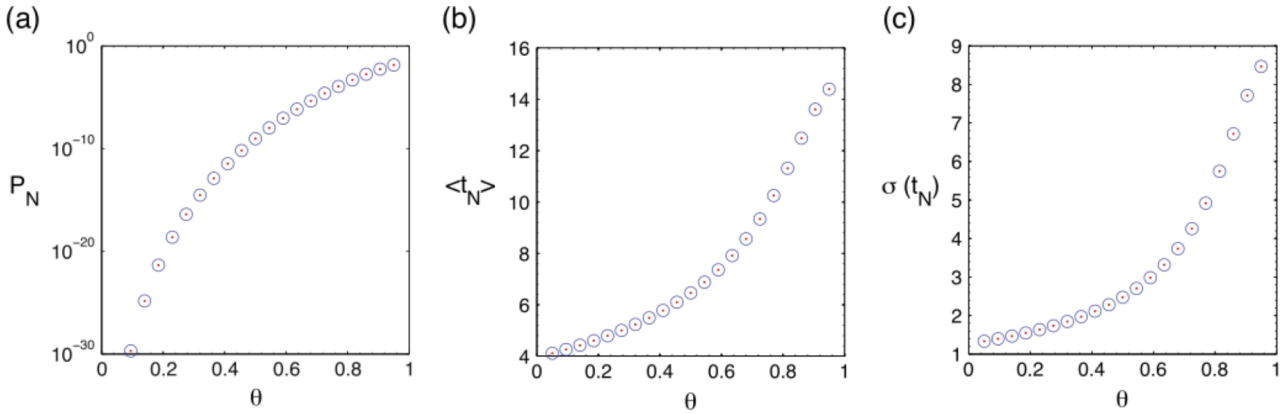

Figure 1. Direct comparison of the analytic solution of the master equation, equations (26), (29) and (30) (blue circles) with direct numerical integration (red dots). (a), (b), and (c) show respectively the probability of establishment, average colonization time, and standard deviation in the colonization time, all as a function of θ for $N = 30$.

colonization trajectories are effectively more deterministic (i.e. less noisy) for lower fitness cells than for higher fitness cells. This is consistent with the preceding discussion, in that the statistically dominant growth trajectories of low-fitness cells are strongly biased to repetitive birth events, and will thus have a low degree of statistical variation.

In the limit of extremely low fitness, $\theta \to 0$, it is relatively straightforward to calculate the entire distribution of colonization times for $N \gg 1$, through a general solution of the generating equation (20). One finds that the colonization times are distributed according to the Gumbel distribution (well-known in extreme value statistics), centred at $t^* = \log N$ (which, according to equation (33), is also the average colonization time in this limit). Explicitly, one finds for the distribution function of colonization times:

$$P(t_N) = \exp\left(-\left(t_N - t^*\right) - e^{-\left(t_N - t^*\right)}\right). \tag{36}$$

This result suggests that the colonization time is drawn as an extreme value in samples of random variables with exponential distribution [24].

3.4. Numerical verification and stochastic realizations

Despite the existence of exact solutions and asymptotic analysis, given the counter-intuitive nature of the results it is reassuring to check them against a numerical integration of the original master equation (1) with the boundary conditions equations (2)–(4). Integration was performed with a 2nd order Runge–Kutta scheme, and equations (6)–(9) were used to evaluate the statistics of interest. In figure 1 we compare the numerical solution to the exact results for the probability of establishment (26), the average time of colonization (29), and the standard deviation of the time of colonization (30). In each case we have perfect agreement as expected. We also compare a numerical evaluation of the exact results with the asymptotic analysis (for large N) in figure 2 for the average time of establishment (33) and the standard deviation of the time of colonization (35), and agreement is found, improving as N increases.

Numerical generation of stochastic realizations of the birth-death process are useful to get better intuition for the actual realizations of successful colonization events. This was achieved by implementing the Gillespie method [25] to generate many realizations of the birth/death stochastic process corresponding to the master equation (1). Results for the mean and standard deviation of establishment times are supported by these numerical results as shown in figure 3. Furthermore, we show some representative realizations of colonization for cells of two different fitnesses (figure 4) demonstrating explicitly that successful events arising from lower-fitness progenitors are indeed typically more explosive and deterministic in their dynamics than those for higher-fitness cells.

3.5. Dualities

Given that our calculations for temporal statistics are valid for any value of θ, it is interesting to compare our results for 'rare dynamics' (i.e. rare successful trajectories arising from cells with $\theta < 1$), with results for 'rare genotypes' (i.e. successful trajectories arising from rare cells with $\theta > 1$). So, we consider $\theta > 1$ in the exact expressions for the growth statistics given in equations (26), (29) and (30). Taking the limit $N \gg 1$ we find that

$$P_N = 1 - \frac{1}{\theta} + O\left(\theta^{-N}\right). \tag{37}$$

Thus for colonization from high-fitness cells, the fraction of successful colonies is of order unity as one would expect. Analysis of equations (29) and (30) shows the exact mapping:

$$\left\langle t_N \right\rangle\big|_{\theta>1}(\theta) = (1/\theta)\left\langle t_N \right\rangle\big|_{\theta<1}(1/\theta), \tag{38}$$

$$\mathrm{var}(t_N)\big|_{\theta>1}(\theta) = \left(1/\theta^2\right)\mathrm{var}(t_N)\big|_{\theta<1}(1/\theta). \tag{39}$$

For $N \gg 1$, using the above relations with the asymptotic results for the temporal statistics found in the low-fitness regime, we have for the high-fitness colonization-time

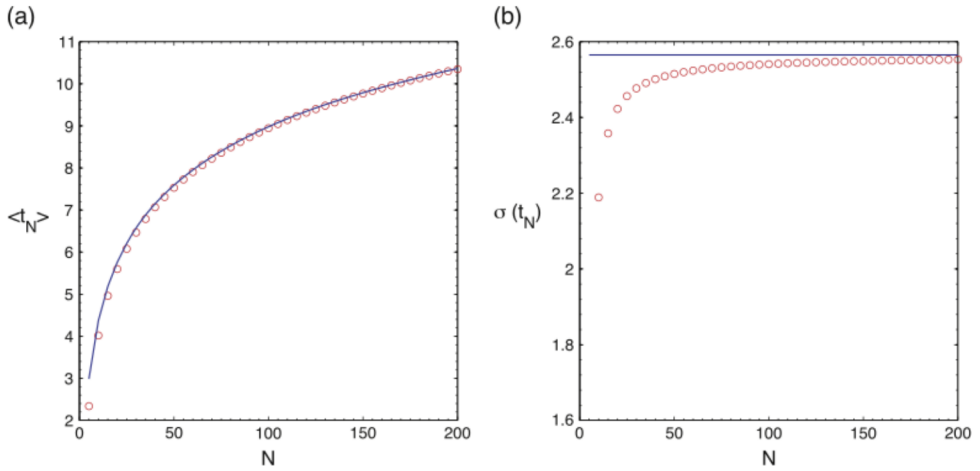

Figure 2. Direct comparison of the analytic solution of the master equation, equations (29) and (30) (red circles) with the asymptotic formulae for large N, equations (33) and (35) (blue lines). (a), (b) show respectively the average colonization time and the standard deviation in the colonization time, both as a function of N, evaluated for $\theta = 0.5$. Convergence of agreement is clear as N increases.

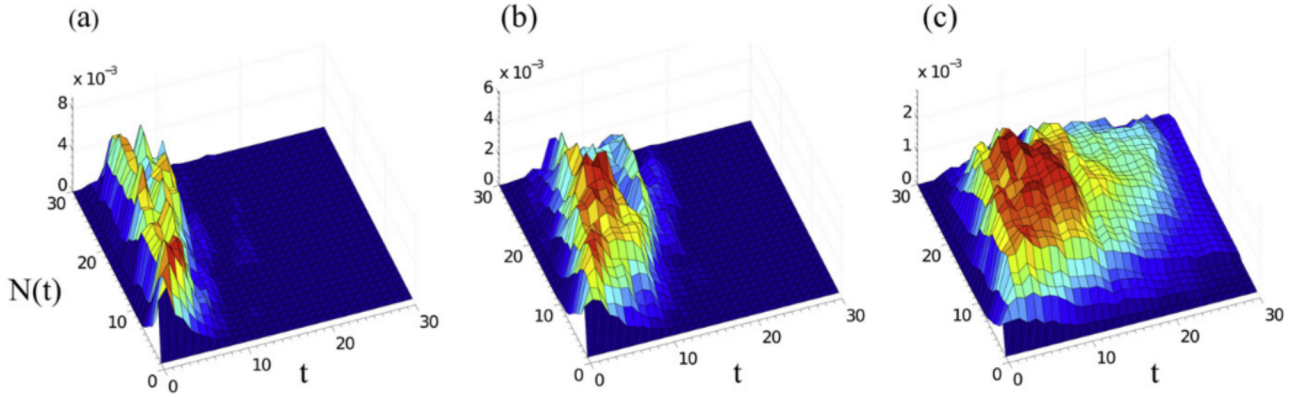

Figure 3. Histograms of successful colonization trajectories. One hundred successful realizations are generated from stochastic simulations of the colonization process for $N = 30$, and three values of cell fitness: (a) $\theta = 0.5$, (b) $\theta = 0.7$ and (c) $\theta = 0.9$. Counting the total number of visits to each point (n, t) gives a normalized two-dimensional histogram, indicating the probability distribution of stochastic trajectories. The vertical axis represents the likelihood that a trajectory will hit the state n at time t. The surface crest denotes the most likely trajectory while the spread indicates the size of statistical fluctuations in the ensemble of trajectories. As fitness increases, typical trajectories take longer to reach the target, and histograms become broader indicating less deterministic dynamics.

statistics:

$$\langle t_N \rangle = \frac{\log N}{(\theta - 1)} + \frac{(\gamma + \log(1 - 1/\theta))}{(\theta - 1)} + O(1/N), \quad (40)$$

$$\mathrm{var}(t_N) = \frac{\zeta(2)}{(\theta - 1)^2} + O(1/N). \quad (41)$$

Interestingly, comparing the leading terms in these expressions with those for the low-fitness colonies, equations (33) and (35), we can see that the expressions are identical if $g_{\mathrm{eff}} = (1 - \theta)$ for $\theta < 1$ colonies is replaced by $g_{\mathrm{eff}} = (\theta - 1)$ for $\theta > 1$. The two solutions are 'dual' to one another for large N: the growth of a rare successful low-fitness colony, of fitness $\theta < 1$, is statistically indistinguishable from the growth of a high-fitness colony of fitness $(2 - \theta) > 1$. This duality is illustrated in figure 5 where the exact expressions for mean and standard deviation of colonization times are plotted as a function of θ for $\theta \in (0, 2)$.

Relating this mathematical result back to the biological context of metastatic colonization leads to the remarkable conclusion: the dynamic progression of micro-metastases at early stages for rare dynamics and rare genotypes are *indistinguishable*. If one were to observe, in vivo, the early exponential growth of a micrometastasis, our results show that one would not be able to infer whether that event was seeded by a pre-adapted or non-pre-adapted cell. A distinction does occur for colony growth beyond the critical size N. Such colonies seeded by low-fitness cells are highly unlikely to sustain the rapid subsequent growth, and would enter a quasi-stable state in which adaptation and microenvironment reconfiguration will presumably enable further, and possibly slow, growth. By contrast, in colonies seeded by fit cells the critical size N is of little relevance, and exponential growth would presumably continue until access to nutrients becomes a limiting factor.

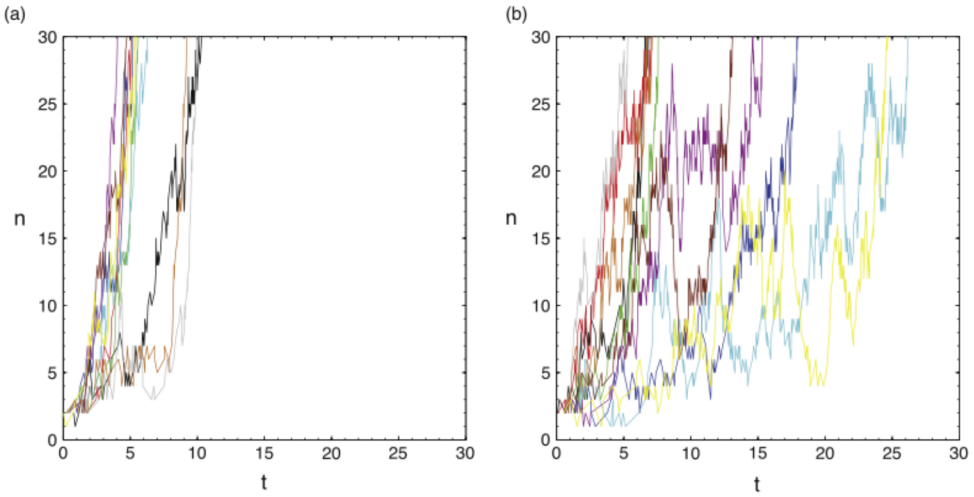

Figure 4. Representative realizations of successful colonization growth for a target tumour size of $N = 30$ cells. The first ten successful trajectories are shown for two values of fitness: (a) $\theta = 0.5$ and (b) $\theta = 0.9$. Note the more rapid dynamics and small variation in the lower fitness trajectories.

A duality of time-inversion is also implied in the exponential growth dynamics of non-pre-adapted cells given in (34). An initial tumour of size N comprising cells with fitness $\theta < 1$ would most likely decay exponentially, with an effective decay rate of $1 - \theta$. The dynamics would be the time-reversed dynamics of our rare event solution, a result reminiscent of reversible solutions in the Freidlin–Wentzell theory of rare event dynamics. Yet another duality exists, in fact, if one considers the dynamics of an initial tumour of size N comprising high-fitness cells ($\theta > 1$), and asks for the temporal statistics of the rare event that such a tumour vanishes. The probability of such an event will be exponentially small, and will typically occur following an exponential decay, for the (time-reversed) reasons discussed at length above. This result implies that spontaneous remission of small tumours comprising fit cells, although rare, will occur rapidly and deterministically, an effect similar to the rare event study here, that would be interesting to pursue in future.

3.6. Heterogeneous populations

It is crucial to recognise that the fitness of a cell attempting colonization depends on both the cell phenotype and the nature of the secondary tissue. Clearly we expect significant heterogeneity in the millions of colonization initiation events happening during the progression of cancer, due both to the heterogeneity of cells leaving the primary tumour, and the heterogeneity of secondary tissue environments in which colonization is attempted. During later stages of metastasis genetic instability, adaptation and microevolution will induce cell heterogeneity within single metastatic tumours. Here though we are only concerned with the very early steps of metastatic colonization. Heterogeneity can then be described by introducing a distribution of fitnesses in the population of progenitor cells (i.e. those single DTCs attempting to proliferate to form a new colony). We denote this distribution $P_{\rm het}(\theta)$. Since this distribution involves properties of both DTCs and secondary tissue sites, it describes the attempted events of colonization throughout the body over the time course of metastatic disease.

For larger values of fitness, $P_{\rm het}(\theta)$ will be a rapidly decreasing function of θ because progressively fitter (or better adapted) cells should be progressively more rare in order to account for metastatic inefficiency. In this regard, it is helpful to introduce ϵ as the fraction of pre-adapted cells in the population:

$$\int_1^\infty \mathrm{d}\theta\ P_{\rm het}(\theta) = \epsilon \ll 1, \tag{42}$$

based on the normalization condition

$$\int_0^\infty \mathrm{d}\theta\ P_{\rm het}(\theta) = 1. \tag{43}$$

The overall probability distribution for successful colonization events in the body is then given by the product of $P_N(\theta)$ and $P_{\rm het}(\theta)$. We will denote this distribution by $P_{\rm succ}$:

$$P_{\rm succ}(\theta) = P_N(\theta)P_{\rm het}(\theta) = \frac{(1-\theta)\theta^{N-1}}{\left(1-\theta^N\right)}P_{\rm het}(\theta)\ , \tag{44}$$

where we have used equation (26). It is convenient to rewrite this expression as

$$P_{\rm succ}(\theta) = \frac{\theta^{N-1}}{\Sigma(\theta)}P_{\rm het}(\theta)\ , \tag{45}$$

where $\Sigma(\theta) = 1 + \theta + \ldots + \theta^{N-1}$.

The probability P_N is an increasing function of θ, and the tail of $P_{\rm het}$ decays rapidly for large θ. Hence $P_{\rm succ}$ should have a maximum for some intermediate value $\theta = \theta_0$, corresponding to the fitness of the most abundant cohort of successful metastatic cells. By comparison, cells with $\theta \gg \theta_0$ are too rare in the initial population to have statistical significance in creating micro metastases, and likewise those with $\theta \ll \theta_0$ are

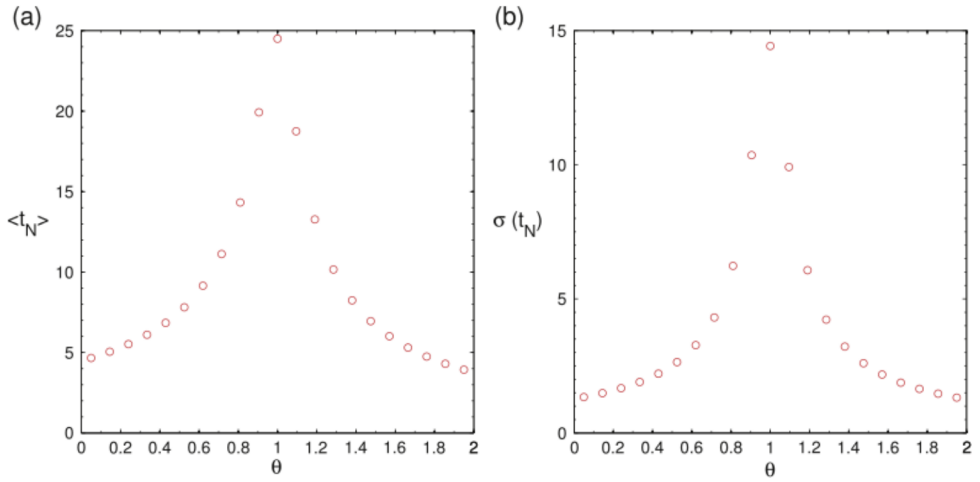

Figure 5. Duality of growth statistics for low and high fitness cells obtained from the exact expressions for the temporal statistics. Average time of colonization (a) and standard deviation in time of colonization (b) are shown for fitness θ varying between 0.0 and 2.0, with the tumour stability threshold $N = 50$. Both the average time and the standard deviation decrease as θ decreases below unity or increases above unity. The duality $\theta \to 2 - \theta$ is exact for $N \to \infty$, but, as can be clearly seen, holds approximately for this moderate value of N.

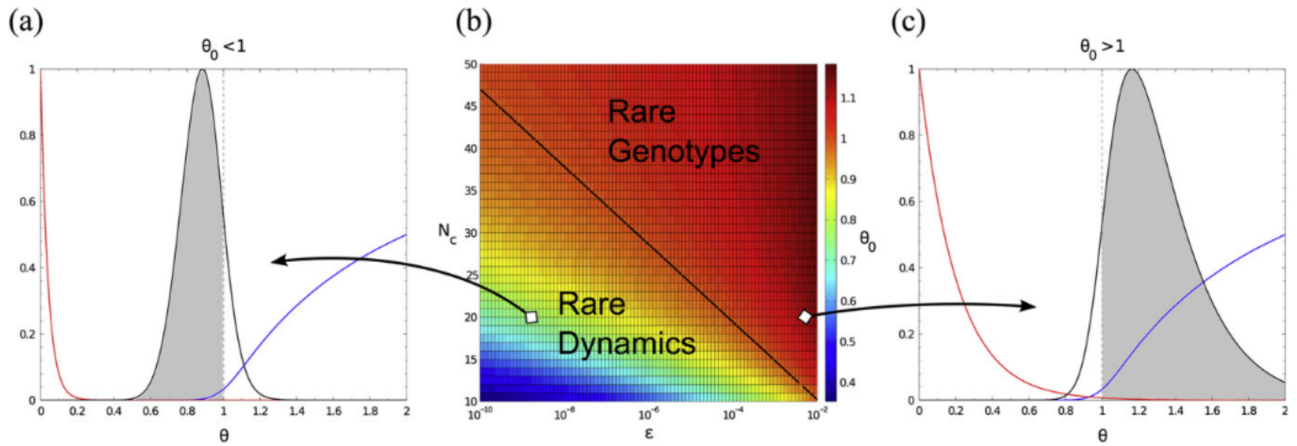

Figure 6. Probability of successful colonization within a heterogeneous population. (a) The net probability of observing a micro-metastasis of size N as a function of fitness θ of the progenitor cell (black curve) is the product of the probability of establishment $P_N(\theta)$ (blue) and the distribution of cell fitness $P_{\text{het}}(\theta)$ (red) in the DTC population, with parameters $N = 20$ and $\epsilon = 2 \times 10^{-9}$. For illustration purposes, the net probability (black curve) is normalized to unity at θ_0. The shaded region emphasises that the majority of successful colonization events arise in this case from rare dynamics. (b) Colour code indicates the values of θ_0 for each combination of critical tumour size N and the fraction of pre-adapted cells ϵ; the black diagonal line corresponds to $\theta_0 = 1$ and indicates the phase boundary separating the domain of rare dynamics ($\theta_0 < 1$) from the domain of rare genotypes, or pre-adapted cells ($\theta_0 > 1$) as the typical mechanism for successful colonization. (c) Same as in (a) but with $N = 20$ and $\epsilon = 5 \times 10^{-3}$. The shaded region emphasises that the majority of successful colonization events arise in this case from rare genotypes. Arrows indicate the location in the parameter space (b) of the two examples (a) and (c).

too unfit. Therefore θ_0 represents the optimal balance between the rarity of initial phenotypes and the rarity of their corresponding dynamics to maximally contribute to nascent metastases. As such, if $\theta_0 < 1$, the colonization process is dominated by rare dynamics (figure 6(a)), whereas if $\theta_0 > 1$ colonization is dominated by rare genotypes (figure 6(c)). We can use the condition $\theta_0 \equiv 1$ to define a phase boundary of the parameter space. This can be written as the condition:

$$\left.\frac{\mathrm{d}P_{\text{succ}}}{\mathrm{d}\theta}\right|_{\theta=1} = 0. \tag{46}$$

Combining equations (45) and (46), we find that the phase boundary condition provides a relationship between the threshold metastasis size N and the distribution of heterogeneity P_{het}:

$$N = 2 \left| \frac{\mathrm{d}}{\mathrm{d}\theta} \log P_{\text{het}}(\theta) \right|_{\theta=1} + 1 \ . \tag{47}$$

To proceed further, it is necessary to make an explicit choice for the form of P_{het}. Unfortunately this distribution is not available experimentally, but one can make a pragmatic choice nonetheless. First, the details of the distribution for low fitness values are not very important since, from equation (47), one will be differentiating the distribution at the threshold fitness value of unity. Second, due to the experimental fact that metastasis is a very inefficient process

[4], a reasonable assumption is that the tail of the distribution decreases exponentially. Such an exponential form could arise, for example, if accumulating discrete random events were required to endow cells with higher fitness values; an idea consistent with popular conceptions of metastatic potential of cells arising from successive mutations of key genes [26].

We therefore take $P_{\rm het}$ to have a generic exponential form:

$$P_{\rm het}(\theta) = A\,\exp\left(-b\theta^{s}\right). \tag{48}$$

The normalization condition (43) fixes $A = sb^{1/s}/\Gamma(1/s)$, where $\Gamma(z)$ is the Gamma function [21]. The second integral condition (42) relates b and s to ϵ as follows

$$\frac{b^{1/s-1}e^{-b}}{\Gamma(1/s)} = \epsilon. \tag{49}$$

Since high-fitness cells are rare, we have $\epsilon \ll 1$. This enables us to give an approximate solution to the transcendental equation (49):

$$b = \log(1/\epsilon). \tag{50}$$

Note this solution is exact if $P_{\rm het}$ is a pure exponential function ($s = 1$).

It is now straightforward to insert the explicit form of $P_{\rm het}(\theta)$ (equation 48) into the condition (47) to derive a relationship between N and ϵ. Taking ($s = 1$) for simplicity we get:

$$N = 2b + 1 = 2\log(1/\epsilon) + 1. \tag{51}$$

This relation demarcates the two domains in the (ϵ, N) parameter space as shown in figure 6(b). Specifically, rare dynamics dominates when the fraction of high-fitness cells is small and the critical size N is not too large. The impact on values of N defining the boundary is only weakly (logarithmically) dependent on ϵ.

3.7. Geometrical considerations

Statistical arguments to follow will allow us to estimate relevant sizes for N. First though, in this subsection, we provide here a simple argument for the lower limit on the size of a nascent metastatic colony which will allow a protective microenvironment for non-surface cells within the cluster. This argument is mainly illustrative, relying purely on geometrical considerations. The true efficacy of a microenvironment depends on many factors such as the biochemical and signalling milieux of the tumour, as well as the morphology of its constituent cells. In addition, for the cluster to be stable, the rate of division of core cells must be high enough to at least replace surface cells which are lost to cell death. This balance will depend on the relative fitnesses of core and surface cells, which is a biological aspect we do not account for in the geometrical arguments below.

Assuming that cells in the metastasis are spherical and of identical sizes, the minimal number of cells required to shield one single core cell is 12, since this is the 'kissing number' for spheres in three dimensions [33]. If indeed a surface of 12 cells provides a sufficiently beneficial microenvironment to the single core cell, one can assign this core cell a high fitness. The subsequent dynamics are, however, not particularly robust, since the progeny of a single high-fitness cell are at significant risk of stochastic extinction. A more robust core would require more than one cell. Using simple geometrical considerations, and assuming a large cluster, one finds that the number of surface cells N_s required to cover N_i internal core cells is given by the relation

$$N_s \sim \left(N_i^{1/3} + 2e^{1/3}\right)^3 - N_i, \tag{52}$$

where e is the packing efficiency, a value that typically varies between $e \sim 0.52$ for randomly packed spheres and $e \sim 0.74$ for Gauss's optimal sphere packing. According to this formula a core comprising 2–10 cells (allowing a more deterministic growth dynamics) would require approximately 25–50 surface cells (figure 7). A more accurate estimate can be gained from numerically packing spheres according to face-centred-cubic close packing, which produces somewhat lower estimates that equation (52). In this case 20-40 surface cells are required to enclose 2–10 core cells. Thus, on geometric grounds, the minimum stability threshold for a proto-metastasis, with greater than one core cell, is around $N \sim 20$–50 cells.

It is possible that the stability threshold could be smaller if surface cells are afforded some degree of protection also. Since cells are likely to adapt their morphology to that of their neighbours, via adhesive interactions, one can think of surface cells in the tumour forming a contiguous surface layer having only one facet in direct contact with the secondary tissue. This partial screening may be sufficient to raise the fitness of surface cells above unity, in which case all cells in the proto-metastasis (both core and surface) would be capable of deterministic growth. In this case, the minimum stability threshold might be as small as approximately 12 cells.

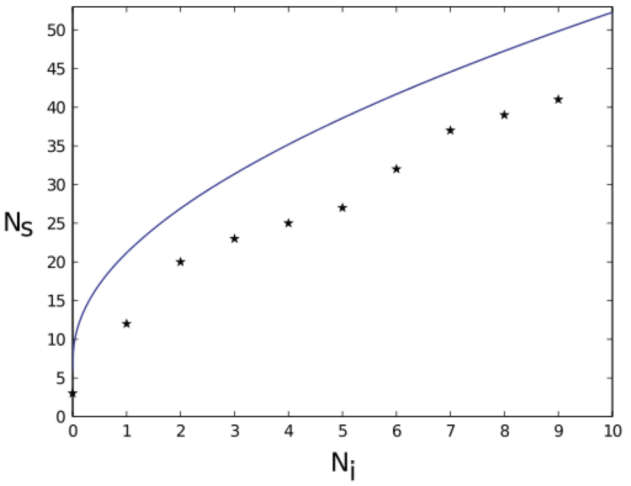

Figure 7. Number of surface cells N_s necessary to shield a core of N_i cells: as given by relationship (52), which overestimates N_s for small tumours (blue line), and by directly counting the number of neighbours in a face-centred-cubic lattice of identical spheres (stars).

3.8. Connections to human cancer

To proceed further, in discussing the relevance of our theoretical results, it is necessary to make closer contact to human cancer, and infer what we can from the dearth of empirical data on the dynamics of metastatic colonization in humans. Indeed, little is known about the properties of DTCs in human cancer. Even measures such as the typical rate of shedding of cancer cells from the primary tumour are poorly known. As such, physiological parameters describing the number and diversity of DTCs will be subject to high uncertainties. Fortunately, for our purposes, it will transpire that only the logarithm of these parameters will be relevant to calibrating our model, so that approximate orders of magnitude in the physiological parameters will suffice.

For concreteness, consider individual DTC colonization attempts occurring over a period of one year in a hypothetical cancer patient who is at risk of metastatic disease, but not in late stage disease from previous successful metastatic colonization. The total number of such attempts, which we denote by M, will be very large, ranging from tens of thousands to hundreds of millions. The typical concentrations of CTCs found in patients with metastasis disease can be estimated at about one to ten cells per millilitre of blood [11, 27], which corresponds to about 10^4–10^5 total CTCs in the human body at any given time. Although a fraction of CTCs will die in the bloodstream, we expect that a significant fraction are rapidly arrested and extravasated from the vascular system as shown by Luzzi *et al* [4] (note, this is an extrapolation from a mouse model), and most of them are not in the blood stream for more than 24 hours [11, 28]. Thus, over a one year period we can expect approximately 10^6–10^7 different DTCs to be attempting colonization. This estimate appears low when compared to other experimental sources. For example, Butler *et al* [29] showed that approximately one million cells per day are shed per one gram of tumour in a rat mammary carcinoma model. If this figure were translated to humans, a one gram tumour would yield over two orders of magnitude greater number of DTCs than quoted above. A lower bound for this number comes from studies of DTCs in bone marrow of prostate tumour patients. In such cases it is estimated that 10^4–10^5 DTCs may be in evidence, most likely in a dormant state [30]. Given this great uncertainty, for our purposes we take M in the range 10^4–10^8.

Assuming that rare dynamics is the dominant mechanism for colonization, we can crudely estimate a range of values for N as follows: let m be the number of colonization attempts in the relevant range $\theta_0 \pm \Delta\theta_0$ where, ultimately, successful events are likely to emerge. It is not possible to estimate systematically what fraction of the M attempts will contribute to m, so given the enormous range of uncertainty in M, we will assume that the range of m is equally large and uncertain. We denote by K the number of m attempts resulting in a metastatic colony. The great inefficiency of metastasis implies that successful colonization events are exceedingly rare [4] and as such K will be orders of magnitude smaller than m. From the definitions given above we have the direct relationship

$$mP_N(\theta_0) \simeq K. \tag{53}$$

Assuming that θ_0 is not too close to unity, equation (31) gives us $P_N(\theta_0) \sim a\theta_0^N$, which together with (53) allows us to express the stability threshold N in terms of the other quantities, viz.

$$N \sim \frac{\log(m/K)}{\log(1/\theta_0)} \sim \frac{\log(m)}{\log(1/\theta_0)}, \tag{54}$$

where we have used in the second step the fact that $m \gg K$. On varying m within the range 10^4–10^8 and θ_0 within the range 0.25–0.85 one finds that N takes a modest range of values, between 10 and 100. The fact that the range of N is relatively well-defined given the enormous uncertainty in the number of attempts and the fitness of DTCs is due to the logarithmic dependence of N on these parameters.

There is a remarkable concurrence of this size range with both geometrical reasoning and biological observations. As discussed in the previous subsection geometrical estimates on the size of a tumour which would enable protection of the core from a hostile microenvironment yields a range for N of about 20–50 cells. Observations of induced metastatic colonization in zebrafish reported that clusters as small as 15–30 cells were able to induce angiogenesis for subsequent growth [31]. There have also been reports of stable metastatic cell clusters in mice of size 30–60 cells [32]. We believe the congruence of biological observations with geometrical considerations and our rare event analysis in identifying 10–100 cells as a threshold size for stability argues for closer experimental examination of such minute clusters, which we term 'proto-metastases'.

3.9. Absolute probabilities and timescales for colonization

We have used equation (54) to determine the possible range of N given crude estimation of physiological ranges of θ_0 and m, finding this range to be 10–100. We can take a different strategy and use (54) to provide a range of values of θ_0 assuming a value for N motivated by biological data, e.g. $N = 30$ as discussed above. Thus, taking $N = 30$ in (54), and the physiological range $m \sim 10^4$–10^8, we find that θ_0 takes values in the rather confined range 0.55–0.75. The absolute probability of colonization for a given event will be trivially given by K/m, and thus will lie in the large range of 10^{-8}–10^{-4}, which simply reflects the large uncertainty in the number of DTCs. Using the range of θ_0 and $N = 30$ we can estimate the average time for a given successful colonization event using the leading term of equation (33). This equation gives the average time in units of the mean time (the inverse of the Poisson rate) of cell death, which we crudely estimate for non-dormant DTCs to be of order one day. One then finds $\langle t_{30} \rangle$ in the range 7.5–13.5 days, i.e. between one and two weeks. This time-scale of initial growth of a successful proto-metastasis is rapid compared to the time interval separating successful events, and relatedly, the typical time-scale of years over which human metastasis clinically progresses.

4. Conclusions

We have presented results arising from a linear birth-death process designed to interrogate the notion that metastatic inefficiency in human cancer may be due to rare dynamical events rather than rare pre-adapted cells. This is a fundamental issue in interpreting and combatting metastatic disease: does colonization of secondary tissue sites arise from deterministic growth of rare pre-adapted cells or from rare stochastic dynamics of common non-pre-adapted cells—the special forces versus infantry in our military metaphor. We have shown, in the context of rare events, that rare dynamics is equally plausible to rare genotypes, and that the dynamics of emerging colonies from each mechanism are, in their early stages, statistically indistinguishable. Despite the paucity of relevant human data, we have been able to provide estimates on key aspects of rare dynamics coloniation—such as the critical colony size N and the mean time of growth—due to the logarithmic dependence of these quantities on the poorly known physiological parameters. Our estimate for N, being in the range 10–100 cells, is, remarkably, in the same range as independent estimates from geometrical considerations and observations of proto-clusters in mice and zebrafish models.

Given the rare event nature of these concepts, direct experimental validation is challenging. The following two experimental avenues might provide such validation. Rare colonization events leading to proto-metastases could occur in proximal stromal tissue surrounding the primary tumour, as well as in distant secondary sites (which has been the focus of the main article). Exhaustive study of high resolution histology sections of such tissue removed along with a primary tumour in a cancer patient would be a means of identifying these stable nascent colonies. If our model is correct one would expect to see a number of tiny metastatic clusters of size 10–100 cells. If only pre-adapted cells are causative agents of metastasis, there will be no significance to this size range, and the size distribution of micro-metastases will show no signature in this range. Better would be high-resolution examination of tissue from secondary sites in a cancer patient with a small recently diagnosed primary tumour. For example, a biopsy from the lung or liver in a breast cancer patient who does not yet have a diagnosis of metastatic disease. However, there are serious medical and ethical concerns about obtaining such tissue samples. An *in vitro* alternative is an experiment inspired by the classic Fidler experiments [6, 8]; to culture a large number of individual cancer cells from an in vivo heterogeneous tumour (preferably from a human patient), but in relatively unfavourable growth conditions. Presumably only a small number of replicates will yield a significant new colony. To determine whether these colonies are due to rare pre-adapted cells or rare dynamics, one would attempt to culture cells from these colonies singly under the same conditions. If a large fraction succeed, then one is observing 'special forces', if a large fraction fail, then one is observing 'infantry'. We must emphasise here that it is imperative that cells are removed from the first round of successful colonies very early on before adaptation has occurred, which can effectively transform 'lucky infantry' into special forces for the given metastatic milieu.

In summary, successful establishment of metastases from low-fitness cells is (i) exponentially rare, (ii) explosive, and (iii) effectively deterministic. As a result of statistical duality, if one were to observe the early dynamics of only successful realizations of colonization arising from low fitness cells, they would resemble perfectly the deterministic, exponential dynamics of pre-adapted cells ($\theta > 1$), leading one to the erroneous conclusion that successful tumours arise from cells with a rare genotype that enabled them to have positive fitness in the secondary tumour site. This duality provides a clear warning in interpretation of metastatic growth statistics—rare dynamics can be as likely as, or more likely than, rare genotypes in causing colonization, and yet can appear phenotypically similar in early stages. Our results are in accord with the seed and soil hypothesis, in that fitness is a key parameter within our model, and fitness, as we have repeatedly stressed, is a co-function of a given DTC (from a given subclone of the primary tumour) in a given secondary tissue environment. The empirical correlation between primary and secondary sites can be interpreted in the context of the rare dynamics model in terms of particular environments providing DTCs with lesser degrees of exponential difficulty of colonization than others. While colonization from rare genotypes provides a rationale for searching for novel drug targets aimed at the particular genetic mutations or phenotypic adaptations that allow such rare cells to be pre-adapted to secondary sites, colonization from rare dynamics, occurring by chance, will require an entirely different paradigm of therapy, presumably more generic than targeted in nature. On the positive side, our results show that the rare dynamics is exponentially sensitive to both the critical cluster size N and the optimal fitness θ_0. Therefore treatments that can increase N and/or lower θ_0, even marginally, will have a significant impact in lowering the overall probability of successful colonization, directly leading to increased periods of an absence of metastatic disease. With particular regard to targeting the critical cluster size, it will be worthwhile to connect the purely demographic analysis in this paper to more detailed cell biological and biomechanical studies of the microenvironment, such as the model of Basan *et al* [18], in which the initial growth of metastases is hampered by homeostatic pressure from the surrounding tissue, essentially rendering all progenitor cells 'unfit', and leading to dynamics analogous to nucleation.

We end with three speculative remarks. First, although this paper has been concerned with metastatic colonization, this entire discussion could be recapitulated in the context of primary tumour initiation. Again, the standard model is rare genotypes (key successive mutations allowing a particular cell to clonally expand), whilst rare dynamics would describe a process of aberrant proliferation against the odds, which very rarely leads to a critical tumour size in which the nascent colony is less fragile due to defining its own new microenvironment. Second, whether one speaks of rare dynamics in terms of metastatic colonization or primary tumour initiation, the rare event statistics show that successful growth will be very rapid. This explosive growth may enable such fragile

colonies to effectively evade immune response. We estimated a time scale of days for growth to a colony size of tens of cells, and it would be interesting to explore this time scale in the context of immune response. Third, and last, our analysis of colonization is not limited, in terms of the modelling, to cancer. Rare dynamics, as a mode for non-pre-adapted individuals to reach critical size in a hostile environment, could find application in infectious disease dynamics within the body, as well as in more traditional areas of modelling such as ecological colonization.

Acknowledgments

The authors thank Barbara Bakker, Jonathan Berg, Paul Davies, William Grady, Barbara Hempstead, Christoph Klein, Dmitry Matyushov, Andrew Rutenberg, Jacob Scott, Kevin Schmidt, D. Eric Smith, Andrew South, Alastair Thompson, Jean Paul Thiery, Thea Tlsty and Cornelis Weijer for discussions. The authors acknowledge support from the National Cancer Institute (Physical Sciences in Oncology Centre Network, U54 CA143682), and TN acknowledges support from the Scottish Universities Life Sciences Alliance.

References

[1] McKinnell R G, Parchment R E, Perantoni A O and Pierce B G 2006 *The Biological Basis of Cancer* 2nd edn (Cambridge New York: Cambridge University Press)
[2] Chaffer C L and Weinberg R A 2011 A perspective on cancer cell metastasis *Science* **331** 1559–64
[3] Weiss L 1990 Metastatic inefficiency *Adv. Cancer Res.* **54** 159–211
[4] Luzzi K J, MacDonald I C, Schmidt E E, Kerkvliet N, Morris V L, Chambers A F and Groom A C 1998 Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases *Am. J. Pathol.* **153** 865–73
[5] Cameron M D, Schmidt E E, Kerkvliet N, Nadkarni K V, Morris V L, Groom A C, Chambers A F and MacDonald I C 2000 Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency *Cancer Res.* **60** 2541–6
[6] Fidler I J 2003 The pathogenesis of cancer metastasis the seed and soil hypothesis revisited *Nat. Rev. Cancer* **3** 453–8
[7] Nowell P C 1976 The clonal evolution of tumor cell populations *Science* **194** 23–28
[8] Fidler I and Kripke M 1977 Metastasis results from preexisting variant cells within a malignant tumor *Science* **197** 893–5
[9] Bernards R and Weinberg R A 2002 A progression puzzle *Nature* **418** 823
[10] Cristofanilli M *et al* 2004 Circulating tumor cells, disease progression, and survival in metastatic breast cancer *New Engl. J. Med.* **351** 781–91
[11] Meng S *et al* 2004 Circulating tumor cells in patients with breast cancer dormancy *Clin. Cancer Res.* **10** 8152–62
[12] Klein C A 2009 Parallel progression of primary tumours and metastases *Nat. Rev. Cancer* **9** 302–12
[13] Chambers A F, Groom A C and MacDonald I C 2002 Dissemination and growth of cancer cells in metastatic sites *Nat. Rev. Cancer* **2** 563–72
[14] van Kampen N G 2007 *Stochastic Processes in Physics and Chemistry* 3rd edn (Amsterdam: Elsevier)
[15] Michor F, Nowak M A and Iwasa Y 2006 Stochastic dynamics of metastasis formation *J. Theor. Biol.* **240** 521–30
[16] Dingli D, Michor F, Antal T and Pacheco J M 2007 The emergence of tumor metastases *Cancer Biol. Ther.* **6** 383–90
[17] Kendal W S 2010 Extinction kinetics for metastatic cancer stem cells *Int. J. Radiat. Biol.* **86** 918–26
[18] Basan M, Risler T, Joanny J-F, Sastre-Garau X and Prost J 2009 Homeostatic competition drives tumor growth and metastasis nucleation *HFSP J.* **3** 265–72
[19] Goel N S and Dyn N 2003 *Stochastic Models in Biology* (Caldwell, NJ: Blackburn Press)
[20] Redner S 2001 *A Guide to First-Passage Processes* (Cambridge: Cambridge University Press)
[21] Abramowitz M and Stegun I A 1965 *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (New York: Dover)
[22] Pendergrass M and Siegrist K 2001 Generalizations of bold play in red and black *Stoch. Process. Appl.* **92** 163–80
[23] Freĭdlin M I and Wentzell A D 1998 *Random Perturbations of Dynamical Systems* 2nd edn vol 260 (New York: Springer)
[24] Gumbel E J 1954 *Statistical Theory of Extreme Values and Some Practical Applications: a Series of Lectures* vol 33 (Washington: US Government Printing Office)
[25] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions *J. Comput. Phys.* **22** 403–34
[26] Yokota J 2000 Tumor progression and metastasis *Carcinogenesis* **21** 497–503
[27] Morgan T M, Lange P H, Porter M P, Lin D W, Ellis W J, Gallaher I S and Vessella R L 2009 Disseminated tumor cells in prostate cancer patients after radical prostatectomy and without evidence of disease predicts biochemical recurrence *Clin. Cancer Res.* **15** 677–83
[28] Patel H, Marer N L, Wharton R Q, Khan Z A, Araia R, Glover C, Henry M M and Allen-Mersh T G 2002 Clearance of circulating tumor cells after excision of primary colorectal cancer *Ann. Surg.* **235** 226
[29] Butler T P and Gullino P M 1975 Quantitation of cell shedding into efferent blood of mammary adenocarcinoma *Cancer Res.* **35** 512–6
[30] Klein C A 2011 Framework models of tumor dormancy from patient-derived observations *Curr. Opin. Genetics Dev.* **21** 42–49
[31] Zhao C, Yang H, Shi H, Wang X, Chen X, Yuan Y, Lin S and Wei Y 2011 Distinct contributions of angiogenesis and vascular co-option during the initiation of primary microtumors and micrometastases *Carcinogenesis* **32** 1143–50
[32] Holmgren L, O'Reilly M S and Folkman J 1995 Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression *Nat. Med.* **1** 149–53
[33] Conway J H and Sloane N J A 1998 *Sphere Packings, Lattices and Groups. A Series of Comprehensive Studies in Mathematics* vol 290 (New York: Springer)