On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell population

Arne Traulsen1, Tom Lenaerts2,3, Jorge M. Pacheco4,5 and David Dingli6

The theory of the clonal origin of cancer states that a tumour arises from one cell that acquires mutation(s) leading to the malignant phenotype. It is the current belief that many of these mutations give a fitness advantage to the mutant population allowing it to expand, eventually leading to disease. However, mutations that lead to such a clonal expansion need not give a fitness advantage and may in fact be neutral—or almost neutral—with respect to fitness. Such mutant clones can be eliminated or expand stochastically, leading to a malignant phenotype (disease). Mutations in haematopoietic stem cells give rise to diseases such as chronic myeloid leukaemia (CML) and paroxysmal nocturnal haemoglobinuria (PNH). Although neutral drift often leads to clonal extinction, disease is still possible, and in this case, it has important implications both for the incidence of disease and for therapy, as it may be more difficult to eliminate neutral mutations with therapy. We illustrate the consequences of such dynamics, using CML and PNH as examples. These considerations have implications for many other tumours as well.

1. Introduction

Tissue homeostasis gives the impression that cell populations are constant and hides the fact that beneath this appearance of stability, there is substantial cell turnover. This is the case in all epithelial tissues (skin, gut, breast) as well as in haematopoiesis. These tissues have a similar architecture where at the root lie tissue-specific stem cells that divide to self-renew and give rise to progeny cells that differentiate into the various types of cells present within the tissue. As long as the dynamic exchanges between these populations are balanced, homeostasis ensues. Therefore, at some macroscopic timescale, the sizes of these hierarchically organized populations, arising from the stem cell population, appear constant.

The stochastic dynamics in such stem cell populations of constant size can conveniently be modelled by the Moran process from population genetics [1–3], a birth–death process (figure 1). Each cell has a certain fitness that is in our case constant. In each time step, a cell is chosen for reproduction at random, but with probability proportional to the fitness. This cell produces identical offspring, replacing a randomly chosen cell, which is considered to differentiate into the next stage. Effectively, this is a death event, because that cell leaves the stem cell compartment and cannot be selected again for reproduction. In this model, cell export captures the initial step in the path towards differentiation of that cell, usually through intermediate cells known as progenitors. Stem cells can acquire mutations, and it is possible that these confer a higher reproductive fitness to the mutated stem cells, hence increasing the chances that the mutant population will invade the whole stem cell compartment. This leads to a so-called clone that corresponds to all the offspring of the
mutant cell in all further differentiation steps. Other mutations may give a fitness disadvantage and such a clone will be eliminated with a probability that increases with the population size and the fitness disadvantage of the cell. As such, these two extremes will induce two potential outcomes: either the mutant population takes over completely or it undergoes extinction. Additionally, certain mutations may not cause a reproductive advantage, i.e. neutral mutations. This case is also a good approximation if selection is weak, i.e. if the product of population size and fitness differences is small [4]. Even in this case, the potential outcomes will, in the long run, be the same as for the two extreme cases mentioned earlier.

However, all organisms have a finite lifetime and so one expects intermediate regimes with mixed populations. Such mixed regimes last longest for neutral mutations when fitness is constant [5]. Such a scenario is not a mere theoretical exercise. Many mutations are indeed neutral [6,7] but the impact of a mutation is cell-context-dependent. In fact, the same mutation may be neutral in one microenvironment but exhibit a fitness difference in another [8–10]. Fitness may also fluctuate in time [11], reflecting changes in the environment of the cells. Here, we do not consider this more complex scenario, but concentrate on fixed fitness values.

There is some evidence, for example, that BCR-ABL expression in a haematopoietic stem cell (HSC) gives no fitness advantage to the cell while the same oncogene expressed in progenitor cells confers to them a fitness advantage [9,10,12–14]. Therefore, the expansion of a BCR-ABL mutant stem cell clone could rely entirely on neutral drift. In a similar manner, neutral drift can explain the incidence, population age structure and spontaneous elimination of PIG-A mutant clones [15,16], a gene required for the synthesis of glycosylphosphatidylinositol-linked proteins that leads to the phenotype observed in paroxysmal nocturnal haemoglobinuria (PNH) [17–19]. In fact, the change of the reproductive fitness of cells owing to acquired mutations is often unrelated to their effect on the entire organism where ‘disease’ may or may not manifest itself as a reduction in the ‘fitness’ of the entire organism.

Let us consider the appearance of a mutation, say BCR-ABL, that has no impact on the reproduction of the HSCs [12–14,20]. Given that the size of the active HSC pool (i.e. the number of HSC that are contributing to haematopoiesis) is small (of the order of 400 cells [21,22]), one expects stochastic effects to have a prominent impact on the clonal evolution of the mutant population [22,23], as also observed in related systems [24]. The theory of the clonal origin of cancer tells us that we start with one cell that arises owing to mutation [25]. How will this clone evolve in time? If the first mutant cell is selected for export, the mutation will disappear from the HSC pool, and only a new mutation will re-initiate such a lineage, an unlikely event [26,27]. By contrast, if such a cell is selected for reproduction, it will generate another BCR-ABL

Figure 1. Possible changes of the number of mutated cells in an elementary time step of the Moran process. (a) The number of mutated cells can increases by one if a wild-type cell differentiates and a mutated cell divides. (b,c) If the same cell type is affected by cell division and differentiation, then the number of mutated cells does not change. (d) The number of mutated cells decreases by one if a mutated cell differentiates and is replaced by a wild-type cell. (Online version in colour.)
expressing cell (because back mutation to a normal state is a very unlikely scenario). As the population of mutant cells increases, the probability that one of them is selected either for reproduction or export also increases. What would be the rate of expansion of such a population? This process is relevant, because persistence of a ‘neutral’ clone can maintain disease in the absence of therapy that can specifically kill such cells [10,19,28]. Moreover, such a process may in part explain why in many tumours, the cancer stem cell population is small, while in others, many of the cancer cells have tumour initiating capability [29,30]. In the following, we provide a mathematical description of such a process and show how stochastic simulations (see appendix) confirm such a behaviour.

2. Stem cell dynamics

Consider a population of constant size \( N \), where cells replicate and differentiate in discrete time steps. In each time step, a cell is selected for reproduction, proportional to fitness and divides symmetrically, producing identical offspring. In the neutral case with \( j \) mutant cells, the probability that we select one of them for reproduction is \( j/N \), while with probability \( (N - j)/N \), we select a normal cell for reproduction; see table 1 for an overview of the mutation. If the mutant cells have a fitness \( r \), the probability of selection is \( rj/(rj + N - j) \). If \( r > 1 \), the mutant has a higher relative fitness compared with the wild-type. If \( r < 1 \), the mutant has a lower relative fitness, whereas if \( r = 1 \), mutant and normal cells have the same fitness. Given that the total population of cells remains constant, then with each reproduction event, we choose a cell for export at random—a typical assumption is that this cell differentiates during cell division and cannot be considered as a primitive stem cell anymore. With probability \( j/N \), we choose a mutant cell for export, whereas with probability \( (N - j)/N \), we choose a normal cell for export. When the process is repeated \( N \) times, on average, each cell would have had one chance to reproduce. This natural timescale of the process is often referred to as generation. For example, in the case of the active HSC pool, we have \( N = 400 \), and each HSC reproduces, on average, about once per year [21,22]. Therefore, when 400 selection–reproduction–export events have occurred, a year will have passed and, on average, each cell would have reproduced once (for neutral mutations).

Note that we have excluded the possibility that a cell divides asymmetrically and produces one differentiated daughter cell and one daughter cell identical to the parent cell. The presence of such asymmetric stem cell divisions would slow down the dynamics between different stem cells. If all cell divisions were asymmetric, the number of stem cells of each type would remain constant, with no room for expansion or extinction of mutated stem cells within the stem cell pool.

2.1. Moran process

The neutral Moran process is a birth–death process with transition probabilities from state \( j \) to state \( k \)

\[
T_{j \rightarrow k} = \begin{cases} \frac{j N - j}{N}, & \text{for } k = j + 1 \\ \frac{1 - j N - j}{N}, & \text{for } k = j \\ 0, & \text{else} \end{cases}
\]

(2.1)

(figure 1). Let us first address the probability that the mutant clone goes extinct. For a neutral mutation, this probability is given by \( 1 - N^{-1} \) in the long run. For short times, we can add the different paths that lead to extinction. The number of such possible paths increases rapidly with \( t \) : there is only one path to extinction in one time step or two time steps, there are two paths to extinction in three time steps, and there are four paths to extinction in four time steps. For example, the probability that the mutant cells go extinct within \( t = 4 \) time steps is given by

\[
M_{1 \rightarrow 0}^t = T_{1 \rightarrow 0} + T_{1 \rightarrow 1} T_{1 \rightarrow 0} + T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} + T_{1 \rightarrow 1} T_{1 \rightarrow 1} T_{1 \rightarrow 0} + T_{1 \rightarrow 1} T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} + T_{1 \rightarrow 2} T_{2 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} + T_{1 \rightarrow 1} T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} = 1 + 2 \frac{1}{2} \left( \frac{N - j}{N} \right)^2 T_{1 \rightarrow 0}
\]

(2.2)

where we have introduced the transition matrix after \( t \) time steps, \( M_{j \rightarrow i}^t = (T_{j \rightarrow i})^t \) for notational convenience. Let us now approximate \( M_{1 \rightarrow 0}^t \) for large \( N \). Because \( j \ll N \), we have

\[
T_{j \rightarrow j+1} = j/N(N-j)/N \approx 1/N \quad \text{and} \quad T_{j \rightarrow j} = 1 - 2j/N \approx 1 - 2(N-j)/N \approx 1 - 2/N \quad \text{for } j \ll N.
\]

The probability to stay in a state is approximately 1, but the probability to move to an adjacent state is close to zero. Thus, we can approximate \( M_{1 \rightarrow 0}^t \) by considering the paths that have up to only a certain number of transitions between states. Taking into account only transitions in which the number of mutants changes only once from one to zero, we obtain

\[
M_{1 \rightarrow 0}^t \approx T_{1 \rightarrow 0} \sum_{k=0}^{j-1} (T_{1 \rightarrow k})^k \approx \frac{1}{N} \sum_{k=0}^{j-1} \left( 1 - \frac{2}{N} \right)^k
\]

(2.3)

where the approximation is valid for large \( N \). For \( t \gg N \), the result can be approximated by \( t/N \). We can improve this approximation by taking into account also those paths in which we reach extinction with up to three transitions between states, which yields for \( t \geq 3 \) the improved approximation

\[
M_{1 \rightarrow 0}^t \approx T_{1 \rightarrow 0} \sum_{k=0}^{j-1} (T_{1 \rightarrow k})^k + T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} \sum_{k=0}^{j-1} \sum_{l=0}^{j-3} (k+1)(T_{2 \rightarrow 1})^l (T_{2 \rightarrow 2})^l.
\]

(2.4)

Going one step further, we could also include the paths involving five transitions, which would lead to two additional terms for \( t \geq 5 \),

\[
T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} \sum_{k=0}^{j-5} \sum_{l=0}^{j-5-k} (k+1) (T_{1 \rightarrow 1})^l (T_{2 \rightarrow 2})^l
\]

\[
+ T_{1 \rightarrow 1} T_{1 \rightarrow 3} T_{3 \rightarrow 2} T_{2 \rightarrow 1} T_{1 \rightarrow 0} \sum_{k=0}^{j-5} \sum_{l=0}^{j-5-k} (k+1) (T_{1 \rightarrow 1})^l (j+1)(T_{2 \rightarrow 2})^l (T_{3 \rightarrow 3})^l.
\]

Note that there are two classes of paths, one in which the state with two mutants is entered and left twice (first line) and one in which the state with three mutants is reached (second line). However, figure 2 illustrates that this approximation is only a marginal improvement over the three-transition approximation equation (2.4), which indicates...
that to derive a better approximation of the time-dependent extinction probability for larger times, we would need to consider a very large number of terms. A similar effect is found for the Wright–Fisher process [31,32].

We are also interested in the conditional average number of mutants \( c \) after \( t \) time steps given that the mutants do not go extinct, \( \langle c \rangle \). This can be obtained by averaging over only those paths that do not lead to state \( k = 0 \),

\[
\langle c \rangle = \frac{\sum_{k=1}^{N} M_1^{t-1} k}{\sum_{k=1}^{N} M_1^{t-1}}. \tag{2.5}
\]

For the numerator, we have at \( t = 1 \)

\[
\sum_{k=1}^{N} M_1^{t-1} k = T_{1-1} \cdot 1 + T_{1-2} \cdot 2 = 1. \tag{2.6}
\]

Thus, the average number of mutants does not change in the first time step. Because the transition probabilities are constant in time, it does not change in the second time step either. This can be iterated to see that the average number of mutants remains constant in time, it does not change in the second time step. This can be iterated to see that the average number of mutants does not change in time—see also the discussion of the branching process in the appendix. Thus, \( \langle c \rangle \) has already been calculated above. With this, we have for \( t < N \)

\[
\langle c \rangle = \frac{1}{1 - M_{1-0}^{t-1}} \approx 1 + \frac{t}{N}. \tag{2.7}
\]

Figure 3 shows that this linear increase is a good approximation not only for small \( t \), but for the first few generations. This is consistent with the arguments above.

Because the conditional fixation time of a single neutral mutant is \( N - 1 \) generations [33–35], we expect that after \( 2N \) generations, the conditional number of mutants is already close to \( N \), as seen in our simulations for \( N = 10 \) in figure 3 (see appendix for an explanation of the simulations).

In order to infer the variance of \( c \), we calculate the quantity

\[
\langle c^2 \rangle = \frac{\sum_{k=1}^{N} M_1^{t-1} k^2}{\sum_{k=1}^{N} M_1^{t-1}}. \tag{2.8}
\]

Again, we apply the same approximation as above and consider only those terms that are based on at most a single transition between states; we find for the numerator of equation (2.8)

\[
\sum_{k=1}^{N} M_1^{t-1} k^2 \approx T_{1-1} \cdot 1 + T_{1-2} \sum_{j=0}^{t-1} (j+1)T_{11} \cdot T_{22}^{t-1-j} \cdot 4. \tag{2.9}
\]
Note that this approximation neglects transitions to states $k > 2$. The next term, in which two transitions are allowed, would lead to two additional terms,

$$T_{1-2}T_{2-1} \sum_{j=0}^{2-j} (j+1)T_{1-1}^{j}T_{2-2}^{2-j} \cdot 1$$

$$+ T_{1-2}T_{2-3} \sum_{j=0}^{2-j} \sum_{i=0}^{2-j-i} T_{1-1}^{i}T_{2-2}^{2-j-i} \cdot 9.$$ (2.10)

However, including this term leads only to a minor improvement of the approximation. Next, we approximate equation (2.9) for large $N$ in the form $\sum_{i=1}^{N} M_{i-1}k^2 \approx 1 + 2k^2$, and find

$$\langle c_i^2 \rangle \approx 1 + \frac{t}{2} + \frac{2t^2}{N}. \quad (2.11)$$

Finally, the error of our quantity $\langle c_i \rangle$ can be approximated by

$$\sqrt{\langle c_i^2 \rangle - \langle c_i \rangle^2} \approx \sqrt{\frac{2t - 1}{N}}.$$ (2.12)

Thus, the leading dependence in $t$ of our approximation is linear, in accord with the numerical results shown in figure 3. Note, however, that our approximation for the error holds quantitatively for very short times only. We note that the limit of the approximation is not the Taylor expansion for large $N$, but the assumption that only few transitions occur. For more precise approximations, a large number of such terms would have to be taken into account.

3. Discussion

The results of our simulations and mathematical analysis have implications for the incidence of acquired HSC disorders. Given that a population with a neutral mutation (or a mutation that confers a very small change in fitness, $N(r - 1) < 1$), on average increases by only one cell per year, then we can conclude that in a disease such as PNH, the original mutation in PIG-A must occur early in life for the clone to reach a threshold compatible with disease (e.g. 10% of the stem cell population being mutated [36]), given that the average age of patients with this disease is in the fourth or fifth decade of life. Of course, this is the average and in principle, growth of the population can be significantly faster—or slower—owing to stochastic effects alone [2]. The other option would be to imply some sort of fitness advantage for the clone either owing to an intrinsic effect, as for instance a second mutation, such as HMGA2 [18, 37], or extrinsic effect such as an immune-mediated attack on normal HSC as proposed by Luzzatto et al. [38]. As discussed elsewhere [26], the probability of a second mutation in the same cell that gives it a fitness advantage is small, and perhaps this is the reason why to date, only a couple of patients have been described with such a second mutation [18]. On the other hand, a fitness advantage could explain instances where the clone is very large, while neutral drift would explain the majority of situations where the clone is large enough to cause disease [19].

In the case of CML, the incidence is almost invariably in people beyond the second or third decade of life. The latent period between the appearance of the Philadelphia chromosome [39] in a HSC and diagnosis is often considered to be in the range of 3–5 years. Therefore, it appears that most mutations leading to the Ph' chromosome occur later in life, presumably because the probability of a chromosomal recombination event leading to such a specific translocation is smaller than the probability of a mutation in PIG-A where a variety of mutations can inactivate the gene [40–42].

Our modelling can also explain an apparent paradox: many mutations in PIG-A inactivate the gene but clinically significant PNH is rare. CML is a significantly more common disorder compared with PNH (by at least one order of magnitude), despite the fact that it is triggered by only a single mutation. Both diseases start with a single...
neutral mutation may be an innocuous passenger but at the level of the stem cell pool, the clone may provide a fitness advantage to their progeny cells [48,49]. The variance in the number of offspring is given by \( \sigma^2 = (0^2 \cdot p + 1^2 \cdot (1 - 2p) + 2^2 \cdot p) - m^2 = 2p \). From the variance, we can obtain the scaling of the expected number of mutated stem cells \( c_j \) given that \( m_0 = 1 \), \( c_j \sim \sigma^2 t / 2 = t / N \) [52]. This scaling holds for large \( t \). Thus, the expected number of mutated stem cells given that they do not go extinct grows by one per generation. This has also been observed in a Moran model of stem cell dynamics [28]. Next, we show that the result can also be obtained directly from the Moran model when the population size is not too small. This illustrates that the conditional growth rate of 1 per generation is valid even for short time \( t \).

A.2. Computer simulations

For the computer simulations, the Moran process was implemented as follows: the system is characterized by the number of mutants \( j \). In each time step, the probabilities to transit to \( j + 1 \) or to stay in \( j \) are calculated based on \( j \), \( N \) and \( r \), see §2. Then, a single pseudorandom number is generated that determines which of the three possible transitions is realized (transfer to \( j + 1 \) or stay in \( j \)). The procedure is then continued in the next state. For the time-dependent extinction probability, the fraction of runs leading to the extinction of all mutants until a certain time is computed (figure 2). For the number of cells given no extinction, as shown in figure 3, only those runs in which extinction did not occur so far are considered.

Appendix

A.1. Branching process

To connect our work to previous papers based on branching processes, we can also consider the dynamics of mutant cells as a branching process [51–53]. As a timescale, we consider a single cell division. With probability \( p = 1/N \), a particular cell produces two offspring cells. With the same probability \( 1 - 2p \), it remains unaffected (and thus can be considered as having a single offspring), the equivalent of self-renewal. Thus, the average number of offspring at time \( t \) is given by \( m_0 = 0 \cdot p + 1 \cdot (1 - 2p) + 2 \cdot p = 1 \), which implies a constant average number of mutated stem cells. Owing to this restriction, we are dealing with a critical branching process [52]. This implies that the mutated stem cells eventually go extinct with probability one. However, because we consider a population of size \( N \), once the mutated cells hit this threshold (i.e. they invade the population) they cannot go extinct anymore. But such a statement is not necessarily helpful in our context, because extinction may take longer than the lifetime of the associated individual. Because extinction is likely, the fact that the average number of mutated stem cells is constant implies large fluctuations of the process [2,52].

As can be seen from this analysis, neutral mutations have many flavours with implications for a variety of disorders that ostensibly arise within the same cell. As in real estate, location is everything. A neutral mutation in some cells may provide a fitness advantage to their progeny cells (BCR-ABL) that can also make them amenable to therapy. A neutral mutation may also be an innocuous passenger but at times leads to a disease that can be as resistant as any other with respect to therapy. Small may not be cute—it may be stubborn, even lethal.

These evolutionary considerations have potential implications for other clonal disorders as well. Genomic sequencing of tumours identifies many mutated genes. The majority of these mutated genes are thought to be ‘passengers’ [50], perhaps implying that they are not important for tumour growth or survival. Such nomenclature may be misleading, as potentially important mutations that are relevant or even critical for the tumour may be ignored.


47. Sirard C et al. 1996 Normal and leukemic SCID-repopulating cells (SRC) coexist in the bone marrow and peripheral blood from CML patients in chronic phase, whereas leukemic SRC are detected in blast crisis. Blood 87, 1539–1548.


