scientific reports

OPEN

Stochastic journeys of cell progenies through compartments and the role of self-renewal, symmetric and asymmetric division

Hanan Dreiwi[1,5], Flavia Feliciangeli[1,2,5], Mario Castro[3], Grant Lythe[1], Carmen Molina-París[1,4] & Martín López-García[1✉]

Division and differentiation events by which cell populations with specific functions are generated often take place as part of a developmental programme, which can be represented by a sequence of compartments. A compartment is the set of cells with common characteristics; sharing, for instance, a spatial location or a phenotype. Differentiation events are transitions from one compartment to the next. Cells may also die or divide. We consider three different types of division events: (i) where both daughter cells inherit the mother's phenotype (self-renewal), (ii) where only one of the daughters changes phenotype (asymmetric division), and (iii) where both daughters change phenotype (symmetric division). The self-renewal probability in each compartment determines whether the progeny of a single cell, moving through the sequence of compartments, is finite or grows without bound. We analyse the progeny stochastic dynamics with probability generating functions. In the case of self-renewal, by following one of the daughters after any division event, we may construct lifelines containing only one cell at any time. We analyse the number of divisions along such lines, and the compartment where lines terminate with a death event. Analysis and numerical simulations are applied to a five-compartment model of the gradual differentiation of hematopoietic stem cells and to a model of thymocyte development: from pre-double positive to single positive (SP) cells with a bifurcation to either SP4 or SP8 in the last compartment of the sequence.

Humans, animals, plants, and even fungi consist of multiple cell types which maintain the organism in homeostasis. All cells share the same DNA but gene expression variation lead to a range of different cell functions. Cell differentiation, controlled through changes in gene expression, enables cells to become more specialised. In embryonic development, non-specialised cells give rise to differentiated and functional cells[1,2]. Sequences of progressive cell-type (or phenotype) changes also take place in adult organisms, leading to cell populations with specific functions. Populations of "stem cells" retain the potential to develop into many different types. Generally speaking, we can say that "stem-like" cells are able to differentiate into multiple cells of a lineage, giving rise to more mature cells. Mathematical models of cell differentiation, fitted to experimental data, have been used to estimate per-cell event rates, to study differentiation timelines, and cellular dysregulation: keratinocyte differentiation and psoriasis pathogenesis[3–5], human gastrointestinal tract cellular differentiation[6,7], blood-cell differentiation and hematopoiesis[8–11], cancer differentiation as in chronic myeloid leukaemia[12], T-cell development and thymic selection[13–15], and the T cell exhaustion process[16–20].

Differentiation involves a series of cell phenotypes, and may involve different spatial locations as well, as in keratinocyte differentiation; the epidermis, the outermost skin layer, is a stratified collection of cells, called keratinocytes, characterised by several differentiated stages. During its lifespan, a keratinocyte transits from a more internal to an external stratum, undergoing biochemical and morphological changes[21]. A similar process is that of colonic cell differentiation; in this case, the endothelial tissue of the human gastrointestinal tract consists of several cell types with different functions, from an outermost layer that acts as a protective barrier and absorbs

[1]School of Mathematics, University of Leeds, Leeds, UK. [2]Systems Pharmacology and Medicine, Bayer AG, Leverkusen, Germany. [3]Grupo Interdisciplinar de Sistemas Complejos (GISC), Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid, Spain. [4]Los Alamos National Laboratory, Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos, NM, USA. [5]These authors contributed equally: Hanan Dreiwi and Flavia Feliciangeli. ✉email: m.lopezgarcia@leeds.ac.uk

nutrients, to an inner one of proliferative stem-like cells. In Ref.[6], the authors investigate cell proliferation and motion in the intestinal crypt (invaginations typical of the colonic tract). Differentiation hierarchies are also found in healthy and cancer systems, where a tumour cell population may arise from cellular mutations. In Ref.[12], a mathematical model of the haematopoietic system representing stem-like cells, progenitors, differentiated and terminally differentiated cells is considered. This hierarchy describes both normal and leukaemic cells (cancer cells) with differences in cellular rates; for instance, leukaemic stem-like cells rapidly divide by self-renewal compared to other differentiated cell types. The authors in Ref.[22] study the plasticity of cancer cells with high and low tumorigenesis potential, making use of a model where cancer cells undergo division to replenish high potential cells or generate cancer cells with lower tumorigenic potential. Cellular de-differentiation also needs to be considered. In chronic infection or cancer, CD8$^+$ T cells (or immune cytotoxic cells) have been observed to acquire dysfunctional (or exhausted) properties[23]. Recent studies have brought a greater understanding of this exhaustion process[20]. Yet, we still dot not completely understand the molecular and epigenetic mechanisms that lead to T cell exhaustion. Unanswered questions include when and at which stage exhaustion becomes irreversible, and to what extent exhausted cells can be re-invigorated.

Advances in genetic labelling[24] have uncovered different types of cell division, which impact the ability of a cell pool to expand or contract. Fate mapping studies reveal linear model structures of cell differentiation, such as the hematopoieitic system[9]. Different division events drive a range of biological processes according to whether a given cell population grows by one new cell (self-renewal), stays the same size (asymmetric division, where one of the daughter cells changes phenotype), or shrinks (symmetric division, where both daughter cells change phenotype)[9]. In such hierarchical systems cells might not be characterised by the same rates throughout the differentiation process[9,25]; an example is found in the hematopoietic system, where a turning point has been observed before the multi-potent progenitor (MPP) stage, where self-renewal is reduced and cell differentiation increased[9]. The heterogeneity of stem-cell populations and their progeny might derive from environmental and intra-cellular perturbations that are still poorly understood. For example, it is unknown whether hematopoietic stem cells (HSCs) undergo symmetric or asymmetric cell division in the unperturbed bone marrow[8–10]. The authors of Ref.[8] found fundamental differences between the normal maintenance of the haematopoietic system, its regulation by challenge, and its re-establishment after transplantation, suggesting that different per-cell rates (of differentiation, death, or division) apply to a given scenario.

Recent experimental techniques have made it possible to track individual cell states (or compartments) and the progeny of a single cell[26,27]. Mathematical and computational approaches have increased our quantitative understanding of cell population dynamics. Deterministic models, which do not incorporate randomness and are typically easier to analyse, may describe the mean dynamics of such populations. However, single-cell behaviour is invariably stochastic. Deterministic compartmental models are used in ecology and cell biology, and in pharmaco-kinetics/pharmaco-dynamics, where they describe the concentration-versus-time curves of a drug following administration and how the drug can influence reaction rates and fluxes. Compartments can represent populations across different scales, from the intra-cellular to the whole-organ level. A single compartment can be thought of as a collection of items, agents, or individuals acting in a homogeneous fashion; agents can differ by form or location, so that a compartment might represent the concentration of a drug in blood or in a given organ.

A multi-compartmental model consists of two or more interconnected compartments and it encodes changes of state (e.g., precursor to product cells) or changes of location. Multi-compartmental models, where compartments are arranged in a sequence or hierarchy, have been used to model cell division and differentiation, from deterministic[28–34] and stochastic[35–38] perspectives. Many of these models have been developed in the context of cancer[30–32,34], and have considered specific hierarchies or parameter ranges (e.g., tumour dynamics with no asymmetric division[30], colorectal crypt[31], multiple mutations[29], age-structured models[34], or Moran-type dynamics[32]). Stochastic models are less frequent. Dingli et al.[36] proposed a mathematical model to illustrate the role of mutations on stem-cell division and the development of tumours. Shahriyari and Komarova[38] proposed a stochastic model for a renewing tissue, addressing the optimisation problem of tissue architecture in the context of mutant production. Clayton *et al.*[35] discussed the classical epidermal proliferation unit (EPU) model of adult epidermal homeostasis, and proposed a single proliferative compartment to better explain experimental observations. Mamis *et al.*[37] considered a three-type branching process to model the dynamics of cell populations in colonic crypts. Refs.[33,39] are recent reviews in this field. The references above differ from the multi-compartmental model analysed here, since we explicitly consider both asymmetric division and cellular de-differentiation. This is not only a generalisation of previous efforts, but it enables us to examine several differentiation processes, such as CD8$^+$ T cell exhaustion reversibility and cancer cell tumorigenesis. Our approach allows the parameters of each cell population (or compartment) to differ, and we do not restrict ourselves to a uni-directional cell flow toward more differentiated states. Thus, we are able to investigate properties caused by the heterogeneity of the different cell populations (or compartments) in a linear-structure model of stem cell differentiation. We also go beyond the more frequent deterministic approach, and analyse the stochastic dynamics of the population with probability generating functions. Finally, by analysing the stochastic dynamics at the single-cell level, we obtain predictions which could be tested (or used for parameter calibration) with novel single-cell experimental methods.

In some circumstances the mean dynamics of cell differentiation might be captured by ordinary differential equations; however, details about cell division, death or differentiation are often better modelled by a stochastic processes, capable of capturing aspects at the single-cell level, genetic and intra-cellular processes, or describing scenarios where a cell population is descended from a few progenitor cells. Within the field of stochastic analysis, the theory of branching processes has been widely used in cell dynamics[40]. For instance, the classic Galton-Watson model, which was originally developed to study the extinction of family surnames[41], has been successfully applied to quantify the progeny of a cell[40]. The theory of branching processes can answer questions related to the limiting behaviour (e.g., probability of extinction or unbounded growth) of cell populations. A natural generalisation is the multi-type branching process, where cells are not all of the same type. These models

can effectively represent cells changing their spatial location[42], or their phenotype[43] over time. They characterise cell dynamics across compartments, where cells in the same compartment follow identical rules. We highlight the seminal work by Matis[44], who proposed a stochastic compartmental model of cell dynamics.

We put forward a general mathematical model with a sequence of N compartments, characterising the process by which, from a stem-cell pool (in compartment C_1), cells undergo differentiation events across adjacent compartments, producing a terminally differentiated progeny population (in compartment C_N). Cells in each compartment can divide, die or transit to adjacent compartments (e.g., representing potentially reversible differentiation or phenotype change). In our stochastic model, differentiation can either be reversible (e.g., cancer cell mutations) or irreversible (e.g., embryonic cell development). Since cells in many tissues are short-lived compared to the life span of the host, a continuous regeneration process is required to maintain cell populations. We consider a close catenary system where all cell populations are explicitly modelled. Any cell generated in the sequence of compartments eventually dies or reaches the final compartment of product cells. However, as we show via a case study in Section "Tracking a thymocyte during its development", alternative compartmental structures (for other potential biological applications) could be considered, for which many of our techniques (e.g., the analysis in Section "Lifeline analysis") can be easily generalised. For example, a cyclic system can be defined if the last compartment is connected to the first one, or a so-called mammillary system, used to represent the blood compartment connected to each organ in the body or cellular states[42].

The paper is organised as follows. In Section "Stochastic compartmental model" we describe the continuous-time Markov dynamics of cells dividing, dying or exiting across a sequence of compartments. The mean behaviour of the system is analysed in Section "Mean number of cells in each compartment". In Section "The progeny of a single progenitor cell", we study the proliferative potential of the system by computing the number of cells in the progeny of a single progenitor cell. In both sections we consider either the situation where differentiation events can be reversible, or a mathematically more simple "irreversible model", where differentiation to the next compartment cannot be reversed. In Section "Lifeline analysis" our focus is a number of summary statistics related to a cellular lifeline which we track over time. In Section "Results", we summarise numerical results inspired by biological applications to illustrate our approach and methods, and examine the impact of asymmetric and symmetric division on the cell population arising from a single progenitor.

Stochastic compartmental model

We propose a stochastic model of cell division, death and differentiation across an ordered sequence of compartments. Cells in a given compartment may represent a common spatial location within the body, or a common phenotype (i.e., representing a biological state, defined by morphology and/or function). In practice, this means that cells in a compartment behave equally, with the same per-cell rates for a given division, death, or differentiation event. We consider a sequence of compartments $C_i, i \in \{1, \ldots, N\}$, which cells follow, behaving independently of each other, as generally assumed in the theory of continuous-time branching processes[40].

Our general stochastic model considers a number of cellular events inspired by some recent mathematical models[3,9,14,42]. We assume that each of these events takes place at a given per-cell rate. Particular situations of interest arise from setting some of the rates equal to zero, so that those events cannot take place, as we illustrate for three different case studies in Section "Results". Each cell in a given compartment, C_i, can divide, die or exit to one of the two adjacent compartments. When a division occurs, both daughter cells might belong to the same compartment as the mother (this event is referred to as *self-renewal*), both daughter cells might instantaneously move to the next compartment (*symmetric division*), or one daughter cell might belong to the same compartment as the mother, and the other to the next compartment (*asymmetric division*)[3,9].

The stochastic model, shown in Fig. 1, is a continuous-time Markov chain (CTMC) $\mathcal{X} = \{(\boldsymbol{C}_1(t), \boldsymbol{C}_2(t), \ldots, \boldsymbol{C}_N(t)) : t \geq 0\}$, where $\boldsymbol{C}_i(t)$ is a random variable that represents the number of cells in compartment C_i at time t, with state space given by $\mathcal{S} = \{0, 1, 2, \ldots\}^N = \mathbb{N}_0^N$. At any given time, the process can be at a particular state $(n_1, \ldots, n_N) \in \mathcal{S}$, so that there are n_i cells in compartment C_i. Thus, cell events, labelled E1 to E5 below, represent transitions between states as follows:

(E1) Self-renewal (cell division where both daughter cells remain in the same compartment as the mother) can occur in any compartment C_i, with per-cell rate λ_i, for $i \in \{1, \ldots, N\}$,
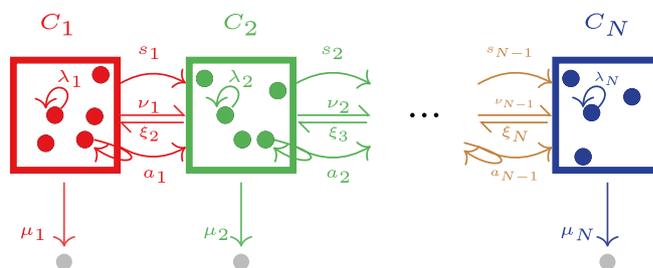

Figure 1. General stochastic model of cell division, death and differentiation for an ordered sequence of compartments. Grey cells represent a death event. Self-renewal events take place with rate λ_i, symmetric division events with rate s_i, and asymmetric division events with rate a_i. Differentiation events happen with per-cell rate ν_i (forward) or ξ_i (backward). Death events have per-cell rate μ_i. Each per-cell rate is a real positive number.

$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{\lambda_i n_i} \to (n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_N).$$

The previous notation indicates that the event moving the process from state $(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N)$ to state $(n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_N)$ happens at rate $\lambda_i n_i$, and represents self-renewal.

(E2) Symmetric division (cell division where both daughter cells instantaneously move to the next compartment) can occur in compartment C_i with per-cell rate s_i, for $i \in \{1, \ldots, N-1\}$,

$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{s_i n_i} \to (n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1} + 2, \ldots, n_N).$$

(E3) Asymmetric division (cell division where one of the daughter cells remains in the same compartment as the mother, while the other goes to the next compartment) can occur in compartment C_i with per-cell rate a_i, for $i \in \{1, \ldots, N-1\}$,

$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{a_i n_i} \to (n_1, \ldots, n_{i-1}, n_i, n_{i+1} + 1, \ldots, n_N).$$

(E4) Differentiation (or migration) between adjacent compartments can occur with per-cell rate ν_i, for $i \in \{1, \ldots, N-1\}$ and ξ_i, for $i \in \{2, \ldots, N\}$,

$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{\nu_i n_i} \to (n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1} + 1, \ldots, n_N),$$
$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{\xi_i n_i} \to (n_1, \ldots, n_{i-1} + 1, n_i - 1, n_{i+1}, \ldots, n_N).$$

(E5) Cells can die in any compartment C_i with per-cell rate μ_i, $i \in \{1, \ldots, N\}$,

$$(n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_N) \to_{\mu_i n_i} \to (n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_N).$$

We assume that cells in the last compartment, C_N, cannot symmetrically or asymmetrically divide, or differentiate to the next compartment. All rates are assumed to be positive real numbers.

Mean number of cells in each compartment

We first study the dynamics of the mean number of cells in each compartment at time t, $\mathbb{E}[C_i(t)]$, described by the following system of ordinary differential equations[44]

$$
\begin{aligned}
\frac{d\,\mathbb{E}[C_1(t)]}{dt} =& -(\mu_1 + \nu_1 + s_1 - \lambda_1)\mathbb{E}[C_1(t)] + \xi_2 \mathbb{E}[C_2(t)], \\
\frac{d\,\mathbb{E}[C_i(t)]}{dt} =& (\nu_{i-1} + a_{i-1} + 2s_{i-1})\,\mathbb{E}[C_{i-1}(t)] - (\mu_i + \nu_i + \xi_i + s_i - \lambda_i)\,\mathbb{E}[C_i(t)] \\
& + \xi_{i+1}\,\mathbb{E}[C_{i+1}(t)], \quad i \in \{2, \ldots, N-1\}, \\
\frac{d\,\mathbb{E}[C_N(t)]}{dt} =& (\nu_{N-1} + a_{N-1} + 2s_{N-1})\,\mathbb{E}[C_{N-1}(t)] - (\mu_N + \xi_N - \lambda_N)\mathbb{E}[C_N(t)],
\end{aligned}
\tag{1}
$$

where $\mathbb{E}[C_i(t)]$ represents the expectation of the random variable $C_i(t)$. These equations constitute a homogeneous first-order linear system of ODEs with constant coefficients, which can be written more succinctly in matrix form as follows

$$\frac{d\mathbf{C}(t)}{dt} = \mathbf{A}\,\mathbf{C}(t), \tag{2}$$

where

$$
\mathbf{C}(t) = \begin{pmatrix} \mathbb{E}[C_1(t)] \\ \mathbb{E}[C_2(t)] \\ \vdots \\ \mathbb{E}[C_{N-1}(t)] \\ \mathbb{E}[C_N(t)] \end{pmatrix}, \quad
\mathbf{A} = \begin{pmatrix}
-\Delta_1 & \xi_2 & 0 & \cdots & 0 \\
\Lambda_1 & -\Delta_2 & \xi_3 & \cdots & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \vdots & \Lambda_{N-2} & -\Delta_{N-1} & \xi_N \\
0 & \cdots & 0 & \Lambda_{N-1} & -\Delta_N
\end{pmatrix},
\tag{3}
$$

and

$$
\begin{aligned}
\Delta_1 =& \mu_1 + \nu_1 + s_1 - \lambda_1, \\
\Delta_i =& \mu_i + \nu_i + s_i + \xi_i - \lambda_i, \quad \Lambda_{i-1} = \nu_{i-1} + a_{i-1} + 2s_{i-1}, \; i \in \{2, \ldots, N-1\}, \\
\Delta_N =& \mu_N + \xi_N - \lambda_N, \quad \Lambda_{N-1} = \nu_{N-1} + a_{N-1} + 2s_{N-1}.
\end{aligned}
$$

The initial value problem of Eq. (2) with $\mathbf{C}_0 = \mathbf{C}(0)$ has a unique solution [[45], Theorem 3.9] given by

$$\mathbf{C}(t) = e^{\mathbf{A}t}\mathbf{C}_0, \tag{4}$$

where $e^{\mathbf{A}t}$ represents the matrix exponential

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \mathbf{A}^2\frac{t^2}{2!} + \mathbf{A}^3\frac{t^3}{3!} + \cdots = \sum_{i=0}^{+\infty} \frac{(\mathbf{A}t)^i}{i!}.$$

The system of equations (2) admits $\lim_{t \to +\infty} \mathbf{C}(t) = \mathbf{0}_N$ (column vector of zeros) as asymptotic solution. It is exponentially stable (solutions of the system for any initial condition converge to zero) if and only if each eigenvalue of $\mathbf{A}$ has a negative real part (see, for example, Ref.[45], Corollaries 3.5 and 3.6]).

We note that, for certain biological applications, some of the rates in Fig. 1 will be zero, and thus, the analysis of such systems would simplify. For instance, differentiation may be irreversible[9,14], so that $\xi_i = 0$ for $i \in \{2, \ldots, N\}$ as shown in Fig. 2. We will refer to this scenario as the *irreversible model*. In this case, and if one considers a single progenitor cell starting in compartment C_1 at time $t = 0$, $\mathbf{C}(0) = (1, 0, \ldots, 0)^T$, it is possible to obtain the mean number of cells for any compartment. In fact, one has

$$\mathbb{E}[C_i(t)] = \begin{cases} e^{-\Delta_1 t}, & i = 1, \\ \left(\prod_{l=1}^{i-1} \Lambda_l\right) \sum_{j=1}^{i} e^{-\Delta_j t} \prod_{\substack{m=1 \\ m \neq j}}^{i} (\Delta_m - \Delta_j)^{-1}, & i \in \{2, \ldots, N\}. \end{cases} \tag{5}$$

Equation (5) is well-defined if $\Delta_i \neq \Delta_j$ for pairs (i, j) with $i \neq j$. If this is not the case, alternative analytic solutions can be found. For example, if $\Delta_i = \Delta_j$ for all $i, j \in \{1, \ldots, N\}$, Eq. (5) simplifies to

$$\mathbb{E}[C_i(t)] = \left(\prod_{l=1}^{i-1} \Lambda_l\right) \frac{t^{i-1}}{(i-1)!} e^{-\Delta_i t}, \quad t \geq 0. \tag{6}$$

We note that this is consistent with existing results in the literature. In particular, the most general solution for the case of repeated eigenvalues can be found in Ref.[46], which analysed the radioactive decay of a chain of nuclides. It is clear that in the irreversible model $\lim_{t \to +\infty} \mathbb{E}[C_i(t)] = 0$, when $\Delta_i > 0 \ \forall i \in \{1, \ldots, N\}$. This is consistent since $\{-\Delta_i : i \in \{1, \ldots, N\}\}$ are the eigenvalues of $\mathbf{A}$ in this case.

For biological applications[3,14] such as those considered in Section "Results", it is of interest to determine the cumulative average number of cells that arrive to the final compartment, C_N, starting with a single "progenitor" or precursor cell in compartment C_1. To this end, in the irreversible model one can set $\lambda_N = \mu_N = 0$, so that cells arriving into C_N accumulate and can be counted. Then, for the last compartment $\Delta_N = 0$ and thus, Eq. (5) leads to

$$\mathbb{E}[C_N(t)] = \left(\prod_{l=1}^{N-1} \Lambda_l\right) \left[\sum_{j=1}^{N-1} e^{-\Delta_j t} \prod_{\substack{m=1 \\ m \neq j}}^{N-1} (\Delta_m - \Delta_j)^{-1} + \prod_{m=1}^{N-1} \Delta_m^{-1} \right], \tag{7}$$

which is well-defined if $\Delta_i \neq \Delta_j$ for all $i, j \in \{1, \ldots, N-1\}$. From the previous equation, we have

$$\lim_{t \to +\infty} \mathbb{E}[C_N(t)] = \prod_{i=1}^{N-1} \frac{\Lambda_i}{\Delta_i}. \tag{8}$$

Interestingly, this limit also holds if $\Delta_i = \Delta_j$ for all $i, j \in \{1, \ldots, N-1\}$ and $\Delta_N = 0$. In this case one can write

$$\mathbb{E}[C_N(t)] = \prod_{i=1}^{N-1} \frac{\Lambda_i}{\Delta_i} - \sum_{j=1}^{N-1} \mathbb{E}[C_j(t)] \prod_{i=j}^{N-1} \frac{\Lambda_i}{\Delta_i}. \tag{9}$$

Under population extinction conditions (that is, when $\Delta_i > 0$ for all $i \in \{1, \ldots, N-1\}$, so that the population of cells in intermediate compartments, $\{1, \ldots, N-1\}$, dies out and only "exiting" cells remain in C_N at late times), $\lim_{t \to +\infty} \mathbb{E}[C_i(t)] = 0$ for $i \in \{1, \ldots, N-1\}$, and thus Eq. (8) also holds.
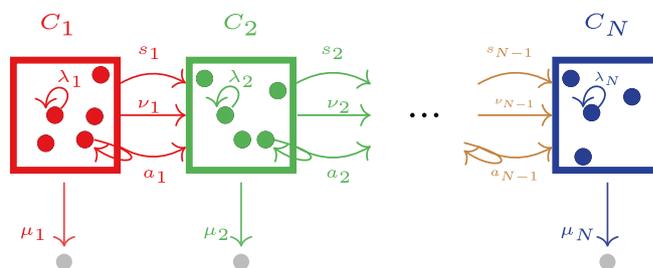

Figure 2. Irreversible stochastic model of cell division, death and forward differentiation for an ordered sequence of compartments. Grey cells represent death events. Self-renewal events occur with per-cell rate λ_i, symmetric division events with rate s_i, and asymmetric division events with rate a_i. Differentiation events happen with per-cell rate ν_i. Death events have (per-cell) rate μ_i. Each per-cell rate is a real positive number.

A particular feature of this system is that cells behave independently from each other. This means that the dynamics of the progeny of a set of M progenitor (or precursor) cells in compartment C_1 at time $t = 0$, can be analysed as M independent stochastic processes. Thus, in Section "The progeny of a single progenitor cell" we consider a number of summary statistics of interest related to the stochastic journeys of cell progenies from a single cell starting at a given compartment (typically compartment C_1).

The progeny of a single progenitor cell

For a cell starting in compartment C_i, we can define G_i to be the random variable representing the total number of cells in the progeny of this cell. Cells in the progeny are the daughters, granddaughters, etc., of the progenitor (or precursor) cell, which originate from division events (either self-renewal, asymmetric or symmetric) in any compartment over time, not including the progenitor cell itself. It is a summary statistic of the process which quantifies the proliferative potential of a single cell in compartment C_i, and its offspring. For example, in Fig. 3, we represent a particular realisation of the stochastic process with $G_1 = 8$.

The mean number of cells in the progeny of a progenitor cell, $m_i = \mathbb{E}[G_i]$, for any initial compartment of interest $i \in \{1, \ldots, N\}$, can be obtained with first-step arguments by conditioning on the next event that occurs in the stochastic process. This approach leads to the following system of equations

$$\Delta_1 m_1 = \Lambda_1 m_2 + 2(\lambda_1 + a_1 + s_1),$$
$$\Delta_i m_i = \Lambda_i m_{i+1} + \xi_i m_{i-1} + 2(\lambda_i + a_i + s_i), \quad i \in \{2, \ldots, N-1\},$$
$$\Delta_N m_N = \xi_N m_{N-1} + 2\lambda_N.$$

The system above can be expressed in matrix form with the column vectors $\mathbf{m} = (m_1, \ldots, m_N)^T$ and $\mathbf{b} = (2(\lambda_1 + a_1 + s_1), \ldots, 2(\lambda_{N-1} + a_{N-1} + s_{N-1}), 2\lambda_N)^T$, as follows

$$\mathbf{J}\,\mathbf{m} = \mathbf{b}, \tag{10}$$

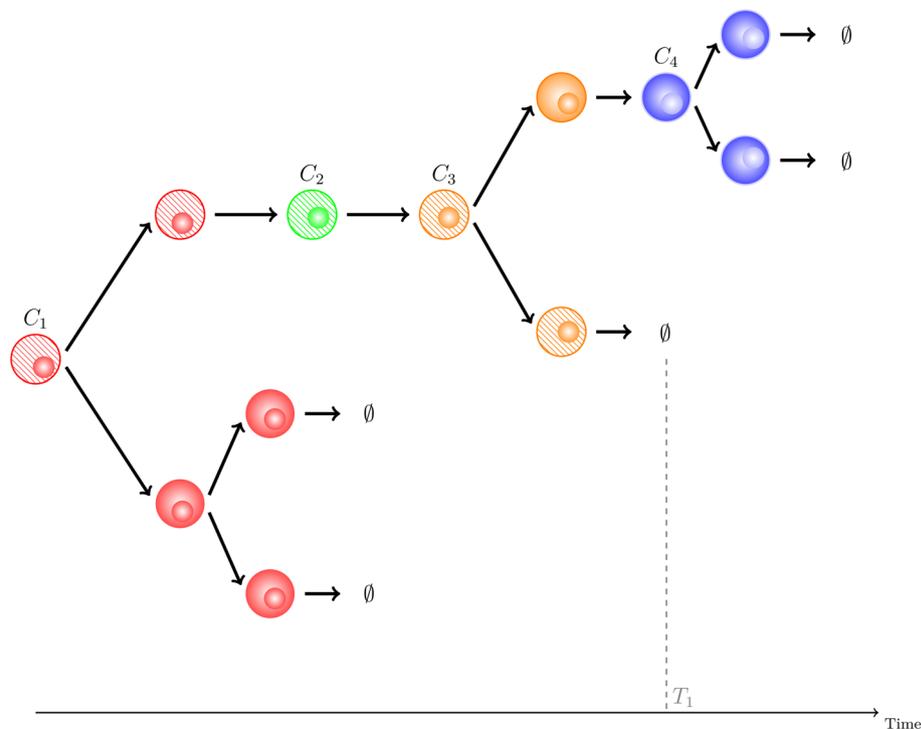with a tri-diagonal coefficient matrix

Figure 3. A realisation of the stochastic process tracking the progeny of a single progenitor cell which starts in compartment C_1. The cell tracked (see Section "Lifeline analysis") is shown as striped. Whenever the tracked cell divides, we follow one of the daughter cells selected at random (see division in C_3). For each cell, the colour indicates the compartment where it is at any given time. Here, the tracked cell dies in C_3 (brown), while its progeny continues up to C_4. In this example, $G_1 = 8 = G_1(1) + G_1(2) + G_1(3) + G_1(4) = 4 + 0 + 2 + 2$.

$$\mathbf{J} = \begin{pmatrix} \Delta_1 & -\Lambda_1 & 0 & 0 & \cdots & 0 \\ -\xi_2 & \Delta_2 & -\Lambda_2 & 0 & \cdots & 0 \\ 0 & -\xi_3 & \Delta_3 & -\Lambda_3 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\xi_{N-1} & \Delta_{N-1} & -\Lambda_{N-1} \\ 0 & \cdots & 0 & 0 & -\xi_N & \Delta_N \end{pmatrix}. \tag{11}$$

One can exploit the tri-diagonal structure of $\mathbf{J}$ to obtain an explicit, or recursive, solution. In particular, by following a Gaussian forward-elimination backward-substitution approach, such as the Thomas algorithm[47,48], one can obtain the recursive equations

$$m_N = \rho_N, \quad m_i = \rho_i - \gamma_i m_{i+1}, \quad i \in \{1, \dots, N-1\}, \tag{12}$$

where $\gamma_1 = -\Delta_1^{-1}\Lambda_1, \rho_1 = 2\Delta_1^{-1}(\lambda_1 + s_1 + a_1)$, and

$$\gamma_i = -\frac{\Lambda_i}{\Delta_i + \xi_i\gamma_{i-1}}, \quad i \in \{2, \dots, N-1\},$$
$$\rho_i = \frac{2(\lambda_i + s_i + a_i) + \xi_i\rho_{i-1}}{\Delta_i + \xi_i\gamma_{i-1}}, \quad i \in \{2, \dots, N\}.$$

Using backward-substitution, this recursive scheme leads to the explicit solution

$$m_i = \sum_{j=i}^{N} (-1)^{j-i} \rho_j \left(\prod_{l=i}^{j-1} \gamma_l\right), \quad i \in \{1, \dots, N\}, \tag{13}$$

where $\prod_{l=i}^{i-1} \gamma_l = 1$. A condition on the parameters arises during the implementation of the recursive scheme

$$\Delta_1 > 0, \quad \Delta_i + \xi_i\gamma_{i-1} > 0, \quad i \in \{2, \dots, N\},$$

so that the mean values $m_1, \dots, m_N$ are finite and non-negative, for all $i \in \{1, \dots, N\}$. This ensures that the number of cells in the progeny of a progenitor cell is finite with probability one, $\mathbb{P}(G_i < +\infty) = 1$, since $m_i = \mathbb{E}[G_i] = \mathbb{E}[G_i|G_i < +\infty]\mathbb{P}(G_i < +\infty) + \mathbb{E}[G_i|G_i = +\infty]\mathbb{P}(G_i = +\infty)$. In the irreversible case, where $\xi_i = 0$ for $i \in \{2, \dots, N\}$, the solution above simplifies to

$$m_i = 2 \sum_{j=i}^{N} \frac{\lambda_j + a_j + s_j}{\Delta_j} \left(\prod_{l=i}^{j-1} \frac{\Lambda_l}{\Delta_l}\right), \quad i \in \{1, \dots, N\}, \tag{14}$$

where for $j = N$, we set $a_N = s_N = 0$. For $i = N$, the empty product above is equal to one, so that $m_N = \frac{2\lambda_N}{\mu_N - \lambda_N}$. For the irreversible model the condition to have finite and non-negative solutions becomes $\Delta_i > 0$ for all $i \in \{1, \dots, N\}$. Direct inspection of Fig. 2 shows that the condition $\Delta_i > 0$ avoids unlimited accumulation of cells in compartment C_i.

Probability generating function
Let us now go beyond the mean number of cells. For the irreversible model we can consider the probability generating function of G_i,

$$\Phi_i(z) = \mathbb{E}(z^{G_i}) = \sum_{k=0}^{+\infty} \mathbb{P}(G_i = k)z^k.$$

The variable G_i counts the cells in the progeny of a progenitor cell starting in compartment C_i, arising from division events (self-renewal, asymmetric and symmetric division), not including the progenitor cell itself. To include the progenitor cell, one can define $S_i \equiv G_i + 1$, and denote the new generating function by $\Psi_i(z) = \mathbb{E}(z^{S_i})$. The total expectation law over all possible first events implies that

$$\mathbb{E}(z^{S_i}) = \sum_{E_j} \mathbb{E}(z^{S_i} \mid \text{event } E_j)\mathbb{P}(E_j), \tag{15}$$

with $E_j \in \{\text{death, differentiation, self-renewal, asymmetric division, symmetric division}\}$. This leads to

$$\Psi_i(z) = \frac{\mu_i}{\Sigma_i}z + \frac{\nu_i}{\Sigma_i}z\Psi_{i+1}(z) + \frac{\lambda_i}{\Sigma_i}z(\Psi_i(z))^2 + \frac{a_i}{\Sigma_i}z\Psi_i(z)\Psi_{i+1}(z) + \frac{s_i}{\Sigma_i}z(\Psi_{i+1}(z))^2, \tag{16}$$

with $\Sigma_i = \mu_i + \nu_i + \lambda_i + a_i + s_i$. Since $S_i = G_i + 1$, one can write $z\Phi_i(z) = \Psi_i(z)$, so that

$$\lambda_i z^2 \Phi_i^2(z) + (a_i z^2 \Phi_{i+1}(z) - \Sigma_i)\Phi_i(z) + \Phi_{i+1}(z)(\nu_i z + s_i z^2 \Phi_{i+1}(z)) + \mu_i = 0. \tag{17}$$

The probability generating functions above agree with the mean values obtained earlier. In particular, by differentiating with respect to z, and setting $z = 1$, we have

$$(\Sigma_i - a_i - 2\lambda_i)\Phi_i'(1) = 2(a_i + \lambda_i + s_i) + (\nu_i + a_i + 2s_i)\Phi_{i+1}'(1), \qquad (18)$$

which can be solved recursively, leading to

$$\mathbb{E}(G_i) = \Phi_i'(1) = 2\sum_{j=i}^{N} \frac{a_j + \lambda_j + s_j}{\Delta_j}\left(\prod_{l=i}^{j-1} \frac{\Lambda_l}{\Delta_l}\right), \qquad (19)$$

in agreement with Eq. (14). We also note that when $i = N$, the cells in the progeny of a single progenitor cell in compartment C_N arise from a linear birth-and-death process, which in discrete time has death probability, $p_d = \frac{\mu_N}{\mu_N + \lambda_N}$, and birth probability, $p_b = \frac{\lambda_N}{\mu_N + \lambda_N}$. Then, a first-step argument for the random variable S_N leads to

$$\Psi_N(z) = p_d z + p_b z \Psi_N^2(z) = z\phi_N(\Psi_N(z)),$$

with $\phi_N(z) = p_d + p_b z^2$. We can then write

$$\Phi_N(z) = \frac{\Psi_N(z)}{z} = \frac{z\,\phi_N(\Psi_N(z))}{z} = \phi_N(z\,\Phi_N(z)),$$

which has solution

$$\Phi_N(z) = p_d\frac{1 - \sqrt{1 - 4x}}{2x} \quad \text{where} \quad x = p_d p_b z^2. \qquad (20)$$

We find $\mathbb{P}(G_N = k)$ as the coefficient of the power of z^k, using the property of the Catalan numbers[49]. In particular, since

$$\sum_{n=0}^{\infty} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x},$$

we obtain

$$\Phi_N(z) = p_d\sum_{n=0}^{\infty} C_n(p_d p_b)^n z^{2n} = \sum_{n=0}^{\infty} \frac{(2n)!}{(n+1)!n!}p_d^{n+1}p_b^n z^{2n} = p_d + z^2 p_d^2 p_b + z^4 2p_d^3 p_b^2 + z^6 5p_d^3 p_b^4 + \cdots.$$

$$(21)$$

Note that $\mathbb{P}(G_N = k+2)/\mathbb{P}(G_N = k) = \frac{4k}{k+3}p_d p_b$. Assuming $\lambda_N < \mu_N$, we verify that $\mathbb{E}(G_N) = \Phi_N'(1) = \frac{2\lambda_N}{\mu_N - \lambda_N}$.

Compartmental analysis of the progeny
Cells in the progeny of a single progenitor can belong to different compartments, as shown in Fig. 3. For a progenitor cell starting in compartment C_i, compartments C_j, $j \in \{1, 2, \ldots, N\}$ with greater proliferative potential will contribute more to G_i. The proliferative potential of compartment j depends on the parameters $\lambda_j, a_j, s_j, \mu_j, \nu_j, \xi_j$, but also on the number of cells arriving into that compartment. It is of interest then to write $G_i = \sum_{j=1}^{N} G_i(j)$, with $G_i(j)$ the number of cells in compartment C_j which belong to the progeny of the progenitor cell from compartment C_i. For example, for the stochastic realisation of Fig. 3, $G_1 = G_1(1) + G_1(2) + G_1(3) + G_1(4) = 4 + 0 + 2 + 2 = 8$.

One can follow similar arguments to the ones in Section "The progeny of a single progenitor cell" to compute the mean quantities $m_i(j) \equiv \mathbb{E}[G_i(j)]$. In particular, for an initial compartment C_i, a first-step argument yields the following equations

$$\Delta_i m_i(i) = 2\lambda_i + a_i + \Lambda_i\, m_{i+1}(i) + \xi_i m_{i-1}(i),$$
$$\Delta_i m_i(i+1) = 2s_i + a_i + \Lambda_i\, m_{i+1}(i+1) + \xi_i m_{i-1}(i+1),$$
$$\Delta_i m_i(j) = \Lambda_i m_{i+1}(j) + \xi_i m_{i-1}(j), \quad j \notin \{i, i+1\},$$

where we implicitly set $\xi_1 = 0$, and $\forall j \in \{1, 2, \ldots, N\}$, $m_{N+1}(j) = m_N(N+1) = m_0(j) = 0$ for notational convenience. Making use of a recursive approach one can show that for any $j \in \{1, \ldots, N\}$, we have

$$m_N(j) = \rho_N(j), \quad m_i(j) = \rho_i(j) - \gamma_i m_{i+1}(j), \quad i \in \{1, \ldots, N-1\},$$

where $\gamma_1 = -\Delta_1^{-1}\Lambda_1$, $\rho_1(j) = \Delta_1^{-1}d_{(1,j)}$, and

$$\gamma_i = -\frac{\Lambda_i}{\Delta_i + \xi_i\gamma_{i-1}}, \qquad i \in \{2, \ldots, N-1\},$$

$$\rho_i(j) = \frac{d_{(i,j)} + \xi_i\rho_{i-1}(j)}{\Delta_i + \xi_i\gamma_{i-1}}, \quad i, j \in \{2, \ldots, N\},$$

with

$$d_{(i,j)} = \begin{cases} 2\lambda_i + a_i, & \text{if } j = i, \\ 2s_i + a_i, & \text{if } j = i+1, \\ 0, & \text{otherwise.} \end{cases}$$

This recursive scheme leads to the solution

$$m_i(j) = \sum_{k=i}^{N} (-1)^{k-i} \rho_k(j) \left(\prod_{p=i}^{k-1} \gamma_p \right), \tag{22}$$

where $\prod_{p=i}^{i-1} \gamma_p = 1$. This expression simplifies further for the irreversible model, since $\xi_i = 0$ for all $i \in \{2, \ldots, N\}$. In this instance, $m_i(j) = 0$ whenever $i > j$, and we can write

$$m_i(i) = \Delta_i^{-1}(2\lambda_i + a_i),$$
$$m_i(j) = \Delta_{j-1}^{-1} \left(\prod_{p=i}^{j-2} \Delta_p^{-1} \Lambda_p \right) \left(d_{(j-1,j)} + d_{(j,j)} \Delta_j^{-1} \Lambda_{j-1} \right), \quad j \in \{i+1, \ldots, N\},$$

for any $i \in \{1, \ldots, N-1\}$. Finally, we note that $\prod_{p=i}^{i-1} \Delta_p^{-1} \Lambda_p = 1$, and $m_N(N) = \frac{2\lambda_N}{\Delta_N}$.

Lifeline analysis

In the previous sections we have analysed several summary statistics of the progeny of a progenitor cell. This section is devoted to summary statistics related to the history of a tracked cell, extending the analysis proposed in Ref.[42], in the situation when all cell divisions are self-renewals; that is, $s_i = a_i = 0$ for all $i \in \{1, \ldots, N\}$. We construct lifelines containing a single cell at any time, starting in compartment C_i. The line continues when that cell divides (by selecting one daughter cell at random) and terminates when the cell dies. In any given compartment, this cell can divide, move to another compartment, or die. Since the only type of division is self-renewal, we use the theoretical "trick" of identifying one of the daughter cells as a continuation of the mother, and the other as a new cell. The dynamics can then be represented by the CTMC $\mathcal{Y} = \{Y(t) : t \geq 0\}$ over the state space $\mathcal{S} = \{C_1, C_2, \ldots, C_N, \emptyset\}$, where $Y(t)$ represents the state of the cell at time t; the cell is either in some compartment, C_j, or it has died (state $\emptyset$). A schematic representation of the process $\mathcal{Y}$ is given in Fig. 4, and a particular realisation of this stochastic process can be seen in Fig. 3, where the tracked cell is shown with stripes.

We study the stochastic process with the following summary statistics:

- The duration, T_i, of the lifeline, starting in compartment C_i,

$$T_i = \inf\{t \geq 0 : Y(t) = \emptyset | Y(0) = C_i\},$$

  which measures the survival potential of cells in the system depending on their initial compartment,
- The number of divisions along a path through the family tree, D_i, which quantifies the proliferative capacity of cells according to their initial compartment, and
- The probability that the lifeline ends in a compartment C_j; that is, $\beta_i(j) = \mathbb{P}(Y(T_i - \Delta t) = C_j)$ for small enough Δt.

Lifespan of a tracked cell
Let $\tau_i = \mathbb{E}[T_i]$ be the mean lifespan of a cell starting in compartment C_i. By conditioning on the events the cell can undergo in the initial compartment, we obtain the following recursive equations

$$(\mu_1 + \nu_1)\tau_1 = \nu_1 \tau_2 + 1,$$
$$(\mu_i + \nu_i + \xi_i)\tau_i = \nu_i \tau_{i+1} + \xi_i \tau_{i-1} + 1, \quad i \in \{1, \ldots, N-1\},$$
$$(\mu_N + \xi_N)\tau_N = \xi_N \tau_{N-1} + 1.$$
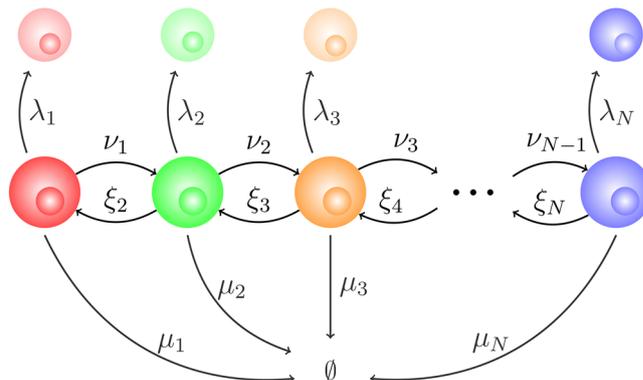
The previous equations have the general solution

Figure 4. Representation of the process $\mathcal{Y}$ to follow the fate of a single cell.

$$\tau_i = \sum_{k=i}^{N} (-1)^{k-i} \bar{\rho}_k \left(\prod_{p=i}^{k-1} \bar{\gamma}_p \right),$$

where $\prod_{p=i}^{i-1} \bar{\gamma}_p = 1$. We have made use of the notation $\bar{\gamma}_1 = -\bar{\Delta}_1^{-1} v_1, \bar{\rho}_1 = \bar{\Delta}_1^{-1}$, with

$$\bar{\gamma}_i = -\frac{v_i}{\bar{\Delta}_i + \xi_i \bar{\gamma}_{i-1}}, \quad i \in \{2,\ldots,N-1\},$$

$$\bar{\rho}_i = \frac{1 + \xi_i \bar{\rho}_{i-1}}{\bar{\Delta}_i + \xi_i \bar{\gamma}_{i-1}}, \quad i \in \{2,\ldots,N\},$$

where $\bar{\Delta}_i = \mu_i + v_i + \xi_i$. A simpler solution is obtained in the irreversible model

$$\tau_i = \sum_{k=i}^{N} \frac{1}{\mu_k + v_k} \left(\prod_{p=i}^{k-1} \frac{v_p}{\mu_p + v_p} \right), \quad i \in \{1,\ldots,N\},$$

where $\prod_{p=i}^{i-1} \frac{v_p}{\mu_p + v_p} = 1$, and we set $v_N = 0$ for notational convenience.

A similar approach allows one to compute the Laplace-Stieltjes transform of T_i, and any of its higher order moments[50]. In particular, one could obtain a system of equations, via first-step arguments, for the Laplace-Stieltjes transform of T_i, $\phi_i(z) = \mathbb{E}[e^{-zT_i}]$. The different order moments can be computed via successive differentiation of the transform, while the probability density function of T_i can be computed via its numerical inversion[51]. For example, in the irreversible model, the second-order moment of the lifespan of a cell starting in compartment C_i is given by

$$\mathbb{E}[T_i^2] = \sum_{j=i}^{N} R_j \left(\prod_{r=i}^{j-1} \frac{v_r}{\mu_r + v_r} \right), \quad i \in \{1,\ldots,N-1\},$$

where $\prod_{r=i}^{i-1} \frac{v_r}{\mu_r + v_r} = 1$, $R_i = \frac{2(v_i \tau_{i+1}+1)}{(\mu_i+v_i)^2}$, and $R_N = \frac{2}{\mu_N^2}$. We note that if the cell starts in the last compartment C_N, $\mathbb{E}[T_N^2] = 2\mu_N^{-2}$, since $T_N \sim \text{Exp}(\mu_N)$.

Number of divisions along a lifeline
Let us define D_i, the number of division events along the history of the tracked cell, starting in compartment C_i. We compute the average value, $\eta_i = \mathbb{E}[D_i]$, with the first-step formula

$$(\lambda_1 + \mu_1 + v_1)\eta_1 = \lambda_1(\eta_1 + 1) + v_1\eta_2,$$
$$(\lambda_i + \mu_i + v_i + \xi_i)\eta_i = \lambda_i(\eta_i + 1) + v_i\eta_{i+1} + \xi_i\eta_{i-1}, \quad i \in \{1,\ldots,N-1\},$$
$$(\lambda_N + \mu_N + \xi_N)\eta_N = \lambda_N(\eta_N + 1) + \xi_N\eta_{N-1}.$$

The previous equations have solution

$$\eta_i = \sum_{k=i}^{N} (-1)^{k-i} \tilde{\rho}_k \left(\prod_{p=i}^{k-1} \bar{\gamma}_p \right), \tag{23}$$

where $\prod_{p=i}^{i-1} \bar{\gamma}_p = 1$, $\tilde{\rho}_1 = \bar{\Delta}_1^{-1} \lambda_1$, and

$$\tilde{\rho}_i = \frac{\lambda_i + \xi_i \tilde{\rho}_{i-1}}{\bar{\Delta}_i + \xi_i \bar{\gamma}_{i-1}}, \quad i \in \{2,\ldots,N\},$$

with $\bar{\Delta}_i$ and $\bar{\gamma}_i$ defined as above. In the irreversible model, this expression simplifies to

$$\eta_i = \sum_{k=i}^{N} \frac{\lambda_k}{\mu_k + v_k} \left(\prod_{p=i}^{k-1} \frac{v_p}{\mu_p + v_p} \right), \quad i \in \{1,\ldots,N-1\},$$

where $\prod_{p=i}^{i-1} \frac{v_p}{\mu_p + v_p} = 1$. We note that in this case $\eta_N = \lambda_N \mu_N^{-1}$, since $D_N \sim \text{Geometric}\left(\frac{\lambda_N}{\mu_N + \lambda_N}\right)$.

A division event may occur at any instant along the lifeline of the cell, which may visit different compartments over time. Thus, one can determine the proliferative potential of the cell during its (eventual) visit to each compartment by considering $D_i = \sum_{j=1}^{N} D_i(j)$, where $D_i(j)$ is the number of divisions which occur exactly in compartment C_j. Average values, $\eta_i(j) \equiv \mathbb{E}[D_i(j)]$, can be computed (again) with first-step arguments. For instance, in the irreversible model, one has

$$\eta_i(i) = \frac{\lambda_i}{\mu_i + \nu_i},$$

$$\eta_i(j) = \frac{\lambda_j}{\mu_j + \nu_j}\left(\prod_{k=i}^{j-1}\frac{\nu_k}{\mu_k + \nu_k}\right), \quad j \geq i+1,$$

for $i \in \{1, \ldots, N\}$ and $\eta_N(N+1) = 0$. We note that the expression above is consistent with the interpretation that, in the irreversible model, $D_i(j) \sim \text{Geometric}\left(\frac{\lambda_j}{\lambda_j + \nu_j + \mu_j}\right)$ restricted to the arrival of the cell to compartment C_j. In general, one can write

$$\eta_i(j) = \mathbb{E}[D_i(j)] = \mathbb{E}[D_i(j) \mid \textit{cell ever visits } C_j]\,\mathbb{P}(\textit{cell ever visits } C_j) + \mathbb{E}[D_i(j) \mid \textit{cell never visits } C_j]\,\mathbb{P}(\textit{cell never visits } C_j),$$

and, since $\mathbb{E}[D_i(j) \mid \textit{cell never visits } C_j] = 0$, the quantity $\eta_i(j)$ accounts for the probability of the cell not visiting this compartment.

More generally, we consider the complete probability distribution of D_i, defined by $\omega_i(n) \equiv \mathbb{P}(D_i = n)$, the probability that a single cell starting in compartment C_i divides exactly n times before it dies or leaves the system, for any non-negative integer n. One can show that

$$\omega_i(n) = \sum_{k=i}^{N}(-1)^{k-i}\hat{\rho}_k(n)\left(\prod_{j=i}^{k-1}\hat{\gamma}_j\right), \quad i \in \{1, \ldots, N\},\ n = 0, 1, 2, 3, \ldots,$$

where $\prod_{j=i}^{i-1}\hat{\gamma}_j = 1$. We have introduced the notation $\hat{\rho}_1(n) = (\lambda_1\omega_1(n-1) + \mu_1 1_{n=0})\hat{\Delta}_1^{-1}, \hat{\gamma}_1 = -\nu_1\hat{\Delta}_1^{-1}$, and

$$\hat{\rho}_i(n) = \frac{\lambda_i\omega_i(n-1) + \mu_i 1_{n=0} + \xi_i\hat{\rho}_{i-1}(n)}{\hat{\Delta}_i + \xi_i\hat{\gamma}_{i-1}}, \quad \hat{\gamma}_i = \frac{-\nu_i}{\hat{\Delta}_i + \xi_i\hat{\gamma}_{i-1}},$$

with $\hat{\Delta}_i = \mu_i + \lambda_i + \xi_i + \nu_i$. We note that $\omega_i(-1) = 0$ and that $1_{n=0}$ is the indicator function, equal to 1 if $n = 0$, and zero otherwise. Thus, the probabilities $\omega_i(n)$ can be computed recursively for increasing values of n, since $\omega_i(n)$ depends on $\omega_i(n-1)$.

Cell death

The probability that the tracked cell dies in compartment j is denoted by

$$\beta_i(j) = \mathbb{P}(Y(T_i - \Delta t) = C_j),$$

for small enough Δt, and for any $i, j \in \{1, \ldots, N\}$. Once again, a first-step argument leads to the following recursive relations

$$(\mu_1 + \nu_1)\beta_1(j) = \nu_1\beta_2(j) + \mu_1 1_{j=1},$$
$$(\mu_i + \nu_i + \xi_i)\beta_i(j) = \nu_i\beta_{i+1}(j) + \xi_i\beta_{i-1}(j) + \mu_i 1_{i=j},$$
$$(\mu_N + \xi_N)\beta_N(j) = \xi_N\beta_{N-1}(j) + \mu_N 1_{j=N},$$

with solution

$$\beta_i(j) = \sum_{k=i}^{N}(-1)^{k-i}\bar{\rho}_k(j)\left(\prod_{p=i}^{k-1}\bar{\gamma}_p\right) \quad i, j \in \{1, \ldots, N\},$$

where $\bar{\rho}_1(j) = \mu_1\bar{\Delta}_1^{-1}1_{j=1}$ for $j \in \{1, \ldots, N\}$, and

$$\bar{\rho}_i(j) = \frac{\mu_i 1_{i=j} + \xi_i\bar{\rho}_{i-1}(j)}{\bar{\Delta}_i + \xi_i\bar{\gamma}_{i-1}}, \quad i \in \{2, \ldots, N\}, j \in \{1, \ldots, N\}.$$

For the irreversible model, for $i \in \{1, \ldots, N\}$, we have

$$\beta_i(i) = \frac{\mu_i}{\mu_i + \nu_i}, \quad \beta_i(j) = \frac{\mu_j}{\mu_j + \nu_j}\prod_{k=i}^{j-1}\frac{\nu_k}{\mu_k + \nu_k}, \quad j \in \{i+1, \ldots, N\}.$$

We conclude this analysis with a comment. A particular advantage of obtaining analytical expressions for the summary statistics is that they allow one to explicitly compute sensitivities, $\partial\beta_i(j)/\partial\theta$, or elasticities, $(\partial\beta_i(j)/\partial\theta)/(\beta_i(j)/\theta)$, with respect to model parameters of interest, θ. This can be rather useful when considering a complex model (with many parameters), as illustrated in Section 'Tracking a thymocyte during its development'. A local sensitivity analysis of this kind provides a quantification of the impact that small perturbations of model parameters can have on a given summary statistics of interest, and is particularly relevant when a subset of the model parameters are being estimated from experimental data sets, and thus, will carry inherent uncertainties. Finally, we refer the reader to the Supplementary Material for extra details on different calculations related to this section.

Results

We illustrate our approach with three case studies; in Section "Asymmetric and symmetric division: the case of four compartments", we implement the methods from Sections "Mean number of cells in each compartment" and "The progeny of a single progenitor cell" to explore the impact of asymmetric and symmetric division events, for the specific case of $N = 4$ compartments. We perform sensitivity analysis for the probabilities of self-renewal, asymmetric and symmetric division. The impact of asymmetric and symmetric division is further analysed in Section "Hematopoietic stem cells: self-renewal, asymmetric and symmetric division", where we consider hematopoietic stem cells, in light of recent experimental data and a mathematical model proposed in Ref.[9]. Finally, in Section "Tracking a thymocyte during its development" we apply the analysis from Section "Lifeline analysis" to an existing model of thymic T cell development[14].

Asymmetric and symmetric division: the case of four compartments

Let us consider the case $N = 4$, for illustrative purposes, where the last compartment, C_4, does not involve any death, division or differentiation events ($\mu_4 = \lambda_4 = \xi_4 = 0$), to represent the terminal accumulation of cells in it. This allows us to quantify the number of cells that *exit* the system formed by the first three compartments, which is of interest in processes such as thymic development[14]. We choose $\mu_i = 1 = \mu$ for all $i \in \{1, 2, 3\}$, so that the unit of time for the system is the mean lifetime of a cell. We want to study the impact of asymmetric and symmetric division on the dynamics, and thus, choose $\nu_1 = \nu_2 = \nu_3 = 1/2$ in the irreversible model, so that all compartments have the same differentiation rates.

Cells can divide in each compartment with per-cell rate ω, and this division represents self-renewal with probability p_{SR}, asymmetric division with probability p_{AD}, and symmetric division with probability p_{SD}. This is equivalent to setting, with the notation introduced in Section "Stochastic compartmental model", $\lambda_i = p_{SR}\omega$, $s_i = p_{SD}\omega$ and $a_i = p_{AD}\omega$, for $i \in \{1, 2, 3\}$. We choose $\omega = 0.9 < 1.0 = \mu$, so that the system has significant proliferative potential, and focus on the following scenarios:

Only-SR. $p_{SR} = 1.0, p_{SD} = 0, p_{AD} = 0.$
Dominant-SR. $p_{SR} = 0.8, p_{SD} = 0.1, p_{AD} = 0.1.$
Dominant-SD. $p_{SR} = 0.1, p_{SD} = 0.8, p_{AD} = 0.1.$
Dominant-AD. $p_{SR} = 0.1, p_{SD} = 0.1, p_{AD} = 0.8.$

Our aim in this section is to explore the impact that asymmetric or symmetric division has on the dynamics of the system (dominant-SR/SD/AD scenarios), compared to the situation where only self-renewal takes place (only-SR scenario). In Fig. 5 we plot the mean number of cells, $\mathbb{E}[C_i(t)]$, in compartments $i \in \{1, 2, 3, 4\}$ for each scenario. For compartments C_i with $i \in \{1, 2, 3\}$, and since $\Delta = \Delta_1 = \Delta_2 = \Delta_3$ and $\Lambda = \Lambda_1 = \Lambda_2 = \Lambda_3$, one can directly make use of Eq. (6). For compartment C_N, $N = 4$, one has

$$\mathbb{E}[C_N(t)] = \left(\frac{\Lambda}{\Delta}\right)^{N-1} - \sum_{j=1}^{N-1} \mathbb{E}[C_j(t)]\left(\frac{\Lambda}{\Delta}\right)^{N-j}.$$
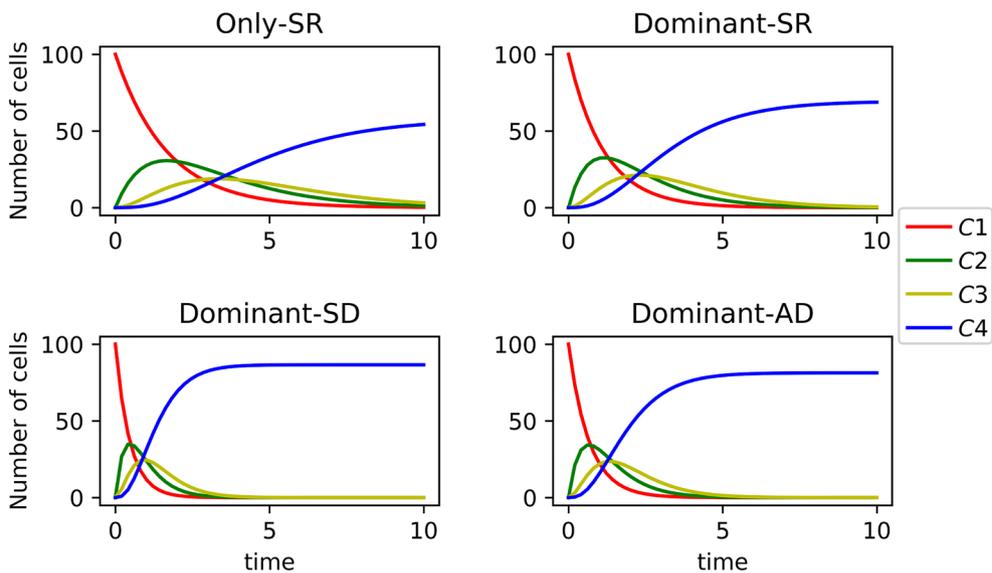

Figure 5. Dynamics of the mean number of cells, $\mathbb{E}[C_i(t)]$, in compartments $i \in \{1, 2, 3, 4\}$ for each of the four scenarios considered in Section "Asymmetric and symmetric division: the case of four compartments".

In Fig. 5, we consider initial conditions $C_1(0) = 10^2, C_2(0) = C_3(0) = C_4(0) = 0$, representing 10^2 initial cells in the first compartment and no cells in the other ones. We observe that an exponential decay in the number of cells in C_1 is followed by sequential increases in the subsequent compartments, until a steady-state number of cells is achieved in the *terminal* compartment, C_4. Interestingly, the dynamics is faster if symmetric or asymmetric division is considered (dominant-SR/SD/AD scenarios compared to only-SR): the decay in compartment C_1 is quicker and the steady-state is reached sooner. The fastest dynamics is observed for the dominant-SD case, where symmetric division is more likely, and the two daughters of a cell move directly to the next compartment. Figure 5 indicates that symmetric or asymmetric division does not only affect the dynamics but also the total mean cell number exiting the system (i.e., reaching the terminal compartment, C_4). This is directly related to the fact that cells can die at any stage in the differentiation process. In particular, $\lim_{t \to +\infty} \mathbb{E}[C_4(t)]$ is significantly larger when asymmetric and (especially) symmetric division can occur. We note that, importantly, the division rate, ω, is equal in all four scenarios. This suggests that, in this type of systems, asymmetric or symmetric division (compared to self-renewal division) facilitates the generation of a larger terminally differentiated population with the same overall *proliferative capacity*; that is, in the only-SR scenario, a larger number of divisions would be required in each compartment for enough cells to escape death and differentiate to the next compartment, and to eventually reach C_4.

Our comments above are consistent with the results shown in Fig. 6, where we plot the mean number of cells, $m_1(j)$, in the progeny of a single cell starting in C_1, belonging to compartments C_j, $j \in \{1, 2, 3, 4\}$, for all four scenarios. In the only-SR scenario, the mean number of cells in the progeny of the progenitor cell decreases monotonically across the sequence of compartments, $m_1(1) > m_1(2) > m_1(3) > m_1(4)$. We note that $m_1(4) = 0$ can be explained since it only accounts for progeny cells which arrive into compartment C_4 as a direct result of cell proliferation, and no symmetric or asymmetric division is considered in the only-SR scenario. We also stress here that this monotonic decrease happens even though the division and differentiation rates are equal in all compartments $j \in \{1, 2, 3\}$. This can be explained by the fact that some cells in the progeny will die before reaching compartments C_2 or C_3. On the other hand, and as discussed in Fig. 5, scenario only-SR leads to the largest mean progeny, $m_1 = m_1(1) + m_1(2) + m_1(3) + m_1(4)$. This suggests that the only-SR scenario is an *inefficient* way to reach a desired population size of terminal cells. Asymmetric and (especially) symmetric division events significantly reduce the number of descendants from a single progenitor cell in all compartments, while maximising the number of terminal (or product) cells (see Fig. 5). In particular, the dominant-SD scenario is characterised by the highest total mean number of terminal cells, $\mathbb{E}[C_4(+\infty)]$, as well as the smallest progeny size, m_1, while leading to the largest progeny, $m_1(4)$, in the terminal compartment.

Finally, the fastest dynamics observed in scenarios with symmetric division, as well as the reduced progeny observed from a single progenitor in C_1, imply that this kind of systems can go from unbounded growth to population extinction at late times, by increasing the number of symmetric division events. We explore this further in the next section, looking at a particular case study.
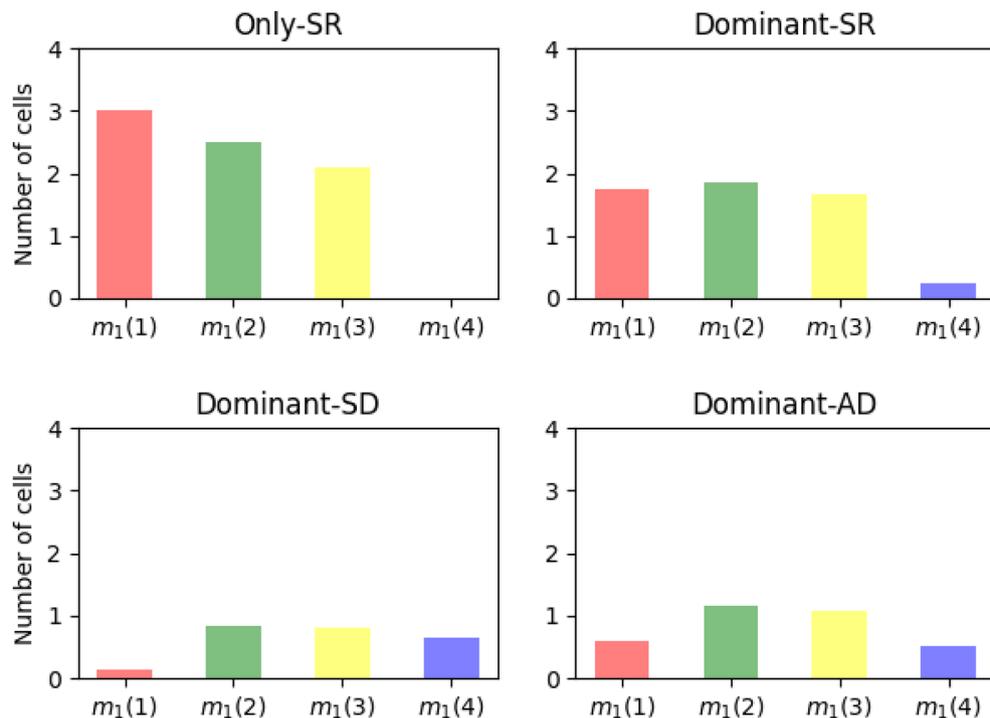

Figure 6. Mean number of cells $m_1(j)$ in the progeny of a single cell starting in C_1, belonging to compartments C_j, $j \in \{1, 2, 3, 4\}$, for all four scenarios.

Hematopoietic stem cells: self-renewal, asymmetric and symmetric division

We consider the model proposed in Ref.[9, Figure 6A] for the gradual differentiation of hematopoietic stem cells (HSCs). HSCs are responsible for the production of all blood cells[52]. In order to maintain such an heterogeneous population, HSCs are able to self-renew and differentiate. Moreover, HSCs need to act continuously and rapidly to either replace short-lived blood cells, or respond to hematopoietic stress arising from events such as bleeding and toxin spotting[53]. Recent advances in flow cytometry and single-cell analysis have shown that HSC cells are a small population compared to the many other blood cell types they can generate. Despite their importance, the molecular mechanisms involved in hematopoietic stem cell maintenance remain unclear.

In order to study simultaneously HSC proliferation and differentiation, Barile et al.[9] propose a novel mathematical model inferred by cell-cycle dependent labelling and HCS fate mapping data. Cells at each state can undergo five different processes: self-renewal, asymmetric cell division, symmetric cell division, direct differentiation and cell death. This leads to the following compartmental sequence: HSC1 → HSC2 → MPP1+MPP2 → MPP3 → HPC1 (see Ref.[9, Figure 6A]) consisting of $N = 5$ different compartments. These represent two stages of hematopoietic stem cells (HSC1 and HSC2), two stages of multi-potent progenitor cells (MPP1 + MPP2 and MPP3), and a final stage of hematopoietic progenitor cells (HPC1).

The model proposed by Barile et al.[9] corresponds to the irreversible model (i.e., $\xi_i = 0$ for $i \in \{2, \ldots, 5\}$) shown in Fig. 2, when we set $N = 5$ compartments and consider the parameter values in Table 1, from Ref.[9, Figure 6]. We note that the "net loss rate" for HSC1 cells, $\Delta_1 = -4.6197 \times 10^{-3}$ days^{-1}, is negative. Thus, the growth of the HSC1 population is unbounded, as can be observed in Ref.[9, Figure S1 H]. This is also shown in Fig. 7, where we simulate the system given by Eq. (2) for the parameters of Table 1.

We note that the parameter calibration performed by Barile et al.[9] predicted relatively negligible symmetric and asymmetric division rates for most of the compartments. Thus, these authors assumed $s_i = a_i = 0$ for all $i \in \{1, \ldots, 5\}$, as given in Table 1 and Fig. 7. We now explore, for this system, the potential role that symmetric and asymmetric division could play. To that end, we perform a sensitivity analysis on the parameters (s_i, a_i), for $i \in \{1, \ldots, 5\}$. In particular, we consider four different scenarios:

- *Only-SR.* All parameters as in Table 1, where $s_i = a_i = 0$ for $i \in \{1, \ldots, 5\}$ and division events correspond to self-renewal, as reported by Barile et al.[9], and shown in Fig. 7.
- *Symm1.* Symmetric division rate $s_1 \in \{10^{-2}, 10^{-1}\}$ for compartment HSC1 is added to the rates in Table 1, with $s_i = 0$ for $i \in \{2, \ldots, 5\}$, and $a_i = 0$ for $i \in \{1, \ldots, 5\}$.
- *SymmAll.* Equal symmetric division rates $s_i \in \{10^{-2}, 10^{-1}\}$ per day, for $i \in \{1, \ldots, 5\}$ are added to the rates in Table 1, with $a_i = 0$ for $i \in \{1, \ldots, 5\}$.

	Δ_i	Λ_i
HSC1	− 0.0046197	0.016497
HSC2	0.0017357	0.007847
MPP1+2	0.0044844	0.032834
MPP3	0.01556	0.16113
HPC1	0.0293	0

Table 1. Parameter values obtained from Ref.[9, Figure 6], where they set $a_i = s_i = 0$ for all $i \in \{1, \ldots, 5\}$. All parameters reported in this section have units of days^{-1}.
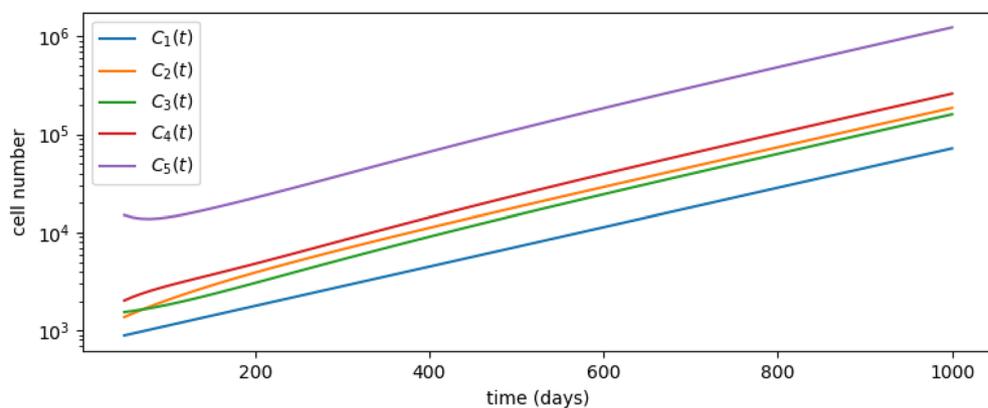

Figure 7. Dynamics of the ODEs system (1) and parameters from Table 1, corresponding to the only-SR scenario. We choose initial conditions $(C_1(0), C_2(0), C_3(0), C_4(0), C_5(0)) = (890, 1370, 1540, 2020, 1.5 \times 10^4)$, taken from Ref.[9, Figure S1 H].

- *AsymmAll.* Equal asymmetric division rates $a_i \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ per day, $i \in \{1, \ldots, 5\}$ are added to the rates in Table 1. Symmetric division only occurs in compartment HSC1, $s_1 = 5 \times 10^{-3} day^{-1}$, chosen so that population extinction is guaranteed.

The net loss rate for compartment i, $\Delta_i \equiv \mu_i + \nu_i + s_i - \lambda_i$, does not depend on the asymmetric division rate, a_i, but it does depend on the symmetric rate, s_i. Thus, just by tuning the symmetric division rate, s_1, of HSC1 cells (Symm1 scenario), we can drastically change the dynamics of the entire system: from unbounded growth (Fig. 7) to population extinction (Fig. 8a). Furthermore, as shown in Fig. 8b, if all the compartments are characterised by a non-zero symmetric division rate (SymmAll scenario), then the resulting population size is smaller but the overall dynamics faster. The change in the asymptotic behaviour of the system, from unbounded growth to extinction, and the smaller number of HSC cells observed in the SymmAll scenario compared to the Symm1 one, can be explained by the transitions across compartments generated by symmetric division events. When a symmetric division event occurs in compartment C_i, two cells in the subsequent compartment C_{i+1} are generated, while the number of cells in compartment C_i decreases by one. Thus, symmetric division events speed up the transition of cells to subsequent compartments and, at the same time, deplete cells from the compartment where the division took place.

In Fig. 9 we explore the impact of increasing the asymmetric division rates, by considering the AsymmAll scenario, where those rates are positive and identical for all compartments. We note that, since asymmetric division rates do not affect the asymptotic qualitative behaviour of the system (i.e., unbounded growth versus extinction), we set $s_1 = 5 \times 10^{-3} day^{-1}$, so that population extinction is guaranteed. When comparing the AsymmAll, Symm1 and SymmAll scenarios, it is clear that increasing the asymmetric division rates leads to a greater number of cells across compartments. This is due to the fact that asymmetric division events increase the number of cells in subsequent compartments without depleting the compartment where the division takes place. In practice, for situations where population extinction is guaranteed at late times (Fig. 9), increasing the asymmetric division rates delays the time when extinction occurs. We note Fig. 9 differs from Fig. 6, but they are not in contradiction. In Fig. 6, the division rate ω is kept constant, while the probability for each type of division event is not, but given by (p_{SR}, p_{SD}, p_{AD}). In Fig. 9, we increase the asymmetric division rate in each compartment instead, effectively increasing the division rate ω, and thus, leading to a greater population size over time in the system.

Studies suggest that very low numbers of HSC cells (HSC1 and HSC2) can maintain a continuous stream of differentiating cells and generate a large number of mature blood cells[8,52]. During hematopoiesis, HSCs cells slowly replace short-lived MPP cells. This heterogeneous population no longer possesses self-renewal ability but
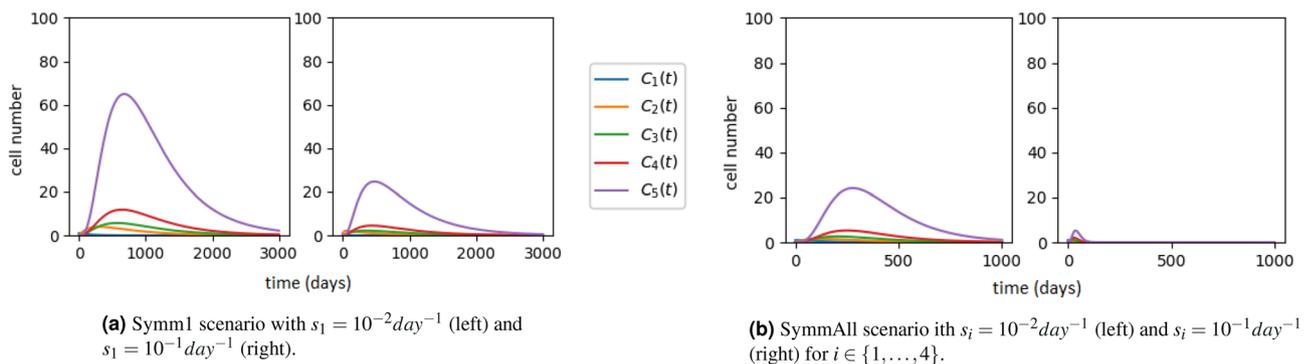

(a) Symm1 scenario with $s_1 = 10^{-2} day^{-1}$ (left) and $s_1 = 10^{-1} day^{-1}$ (right).

(b) SymmAll scenario ith $s_i = 10^{-2} day^{-1}$ (left) and $s_i = 10^{-1} day^{-1}$ (right) for $i \in \{1, \ldots, 4\}$.

Figure 8. Dynamics of the ODE system (1) with initial conditions $(C_1(0), C_2(0), C_3(0), C_4(0), C_5(0)) = (1, 0, 0, 0, 0)$. Parameter values as in Table 1, except for symmetric division rates. (**a**) Scenario Symm1, where the symmetric division rate s_1 is positive only in the HSC1 compartment. (**b**) Scenario SymmAll, where the symmetric division rate s_i is positive and identical for all compartments.
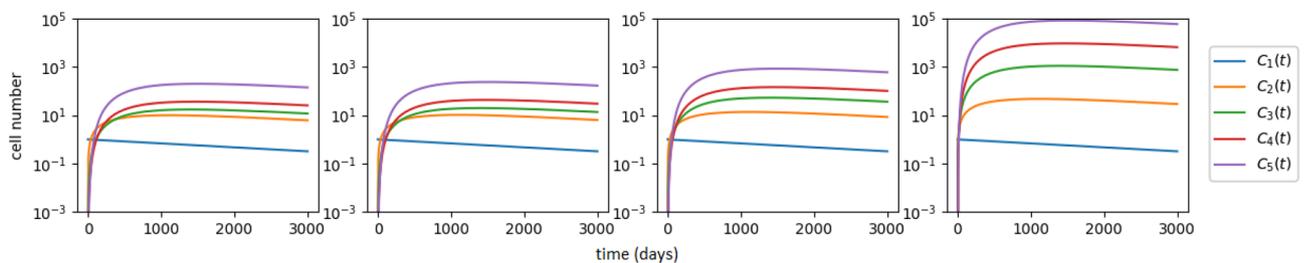

Figure 9. Dynamics of the system (1) and parameter values as in Table 1, except for $s_1 = 5 \times 10^{-3}$ day^{-1} and asymmetric division rates. AsymmAll scenario with equal asymmetric division rates $a_i \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ per day, in all compartments, and initial conditions $(C_1(0), C_2(0), C_3(0), C_4(0), C_5(0)) = (1, 0, 0, 0, 0)$.

still retains differentiation potential[52]. We note here that the parameter values estimated by Barile et al.[9] result in an almost zero net loss rate for HSC2s and MPP1+MPP2 cells; that is, $\Delta_2 \approx \Delta_3 \approx 0$. This agrees with the hypothesis that the self-renewal rate of HSC2 and MPP1+MPP2 cells is sufficient to maintain, alone, the populations of more differentiated cells (e.g., MPP3, HPC1), with minimal input from HSC1 cells. Thus, it is pertinent to study the progeny of a single HSC (HSC1 or HSC2), and study how symmetric and asymmetric division events influence it. To do this, we implement Eq. (14) in the irreversible model for $i \in \{1, 2\}$ and $j \in \{i, \ldots, 5\}$, and compute $m_i(j)$, the mean number of cells within the progeny of a single cell from compartment C_i (HSC1 or HSC2) in subsequent compartments C_j, $j \in \{i, \ldots, 5\}$. The results are shown in Fig. 10, where we plot $m_i(j)$ for the Symm1, SymmAll and AsymmAll scenarios, for $i \in \{1, 2\}$ and $j \in \{i, \ldots, 5\}$. We note that the Symm1 scenario is not considered in Fig. 10 for $i = 2$, since changes in the symmetric division rate, s_1, do not affect $m_2(j)$. First, we observe that the mean number of cells in the progeny of a single HSC1 progenitor across compartments, $m_1(j)$, increases for increasing values of j; that is, for more differentiated cells regardless of the scenario. Indeed, most cells within the progeny of a single HSC1 progenitor belong to the last HPC compartment, consistent with the dynamics observed in Figs. 8 and 9.

The trend of $m_1(j)$ is drastically different for scenario AsymmAll compared to scenarios Symm1 or SymmAll, for increasing values of the corresponding rate (s_1 in Symm1, s_k, $k \in \{1, \ldots, N\}$ in SymmAll, and a_k, $k \in \{1, \ldots, N\}$ in AsymmAll). In the AsymmAll scenario, $m_1(j)$ is an increasing function of the asymmetric division rate a_k, whereas in the Symm1 and SymmAll scenarios, $m_1(j)$ is a decreasing function of the symmetric division rate. This agrees with the dynamics shown in Fig. 8, where increasing values of the symmetric division rate prevent population growth, guaranteeing extinction at late times. An increase of the asymmetric division rate leads to significant production of MPP and HPC cells (see $m_1(4)$ and $m_1(5)$ in Fig. 10) within the progeny, and thus, it could potentially play a role in situations of hematopoietic stress. Similar behaviour can be observed for $m_2(j)$. In agreement with $m_1(j)$ in Fig. 10, we observe a decrease in $m_2(j)$ for scenario SymmAll and an increase in $m_2(j)$ for scenario AsymmAll, as a function of the corresponding division rate. Finally, we compare the Symm1 and SymmAll scenarios. We observe that when the symmetric division rate equals 10^{-2} per day, the number of cells in the progeny of a single HSC1 progenitor differs by almost an order of magnitude between the Symm1 and SymmAll scenarios; that is, symmetric division events taking place in all compartments (SymmAll) lead to smaller progeny from an initial HSC1 cell in subsequent compartments, compared to the Symm1 scenario. When increasing the symmetric division rate to 10^{-1} per day, the difference between the number of cells, $m_1(j)$, within the progeny for the Symm1 and SymmAll scenarios increases further. Clonal hematopoiesis that has been observed in older mice and humans[54] is strictly linked to the accumulation of mutations in the HPC population and an increasing risk of leukemia. Thus, our results suggest that symmetric division could be a possible way to control cell differentiation and limit mutation accumulation in the hematopoietic system.

Tracking a thymocyte during its development

We now consider the T cell thymic development model proposed in Ref.[14, Model 2], and shown in Fig. 11. Double negative (DN) thymocytes differentiate to become pre-selection DP thymocytes (pre-DP). In this model, pre-DP is the first compartment, which will contain an initial number of cells (initial condition, $C_1(0)$). Pre-DPs undergo maturation in the thymus. These cells can progress to the double positive stage (post-DP), where thymocytes express both CD4 and CD8 co-receptors. Post-DP cells that are positively selected transition to the single positive (SP) stage, where they can express either the CD4 or CD8 co-receptor. Some of these cells will then reach the periphery as (single) CD4 or CD8 SP cells.

We exploit this particular model to illustrate the applicability of the analysis developed in Section "Lifeline analysis". This case study also allows us to show how these methods can easily be adapted to different compartment topologies. In this case, we have a compartmental bifurcation, rather than a linear sequence of compartments.

First, it is of interest to estimate the percentage of pre-DP thymocytes that are predicted to die in each of the compartments during development, and the percentage that successfully reach the periphery instead (either as a CD4 or CD8 SP cell). It is clear that our arguments in Section "Lifeline analysis" can easily be adapted to do so. In particular, one can slightly redefine the probabilities $\beta_i(j)$ in Section "Lifeline analysis", with $i = 1$ (i.e., a single pre-DP thymocyte being tracked), as

$$\beta_1(1) = \text{probability that the pre-DP thymocyte dies in the pre-DP compartment,}$$
$$\beta_1(2) = \text{probability that the pre-DP thymocyte dies in the post-DP compartment,}$$
$$\beta_1(4) = \text{probability that the pre-DP thymocyte dies in the CD4 SP compartment,}$$
$$\beta_1(8) = \text{probability that the pre-DP thymocyte dies in the CD8 SP compartment,}$$
$$\beta_1(4P) = \text{probability that the pre-DP thymocyte reaches the periphery as a CD4 SP cell,}$$
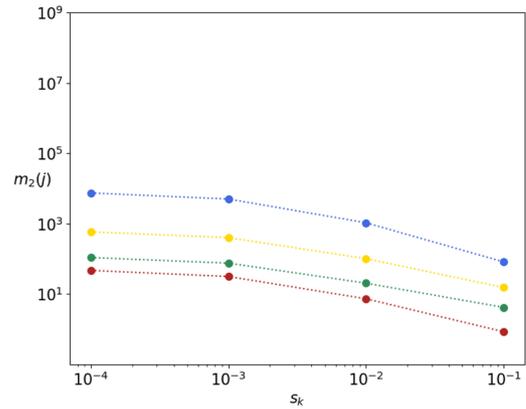$$\beta_1(8P) = \text{probability that the pre-DP thymocyte reaches the periphery as a CD8 SP cell.}$$

Similar solutions to those derived in Section "Lifeline analysis" can be obtained by incorporating the compartmental bifurcation in the first-step analysis, leading to
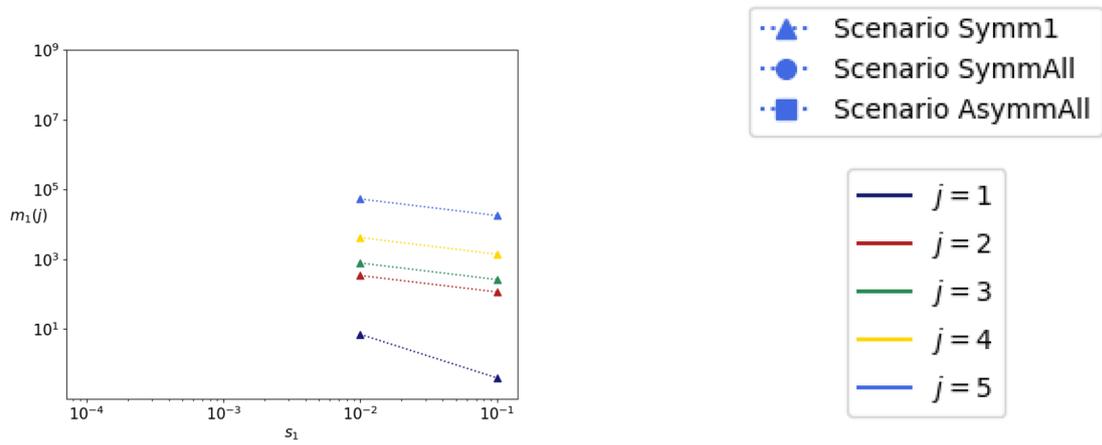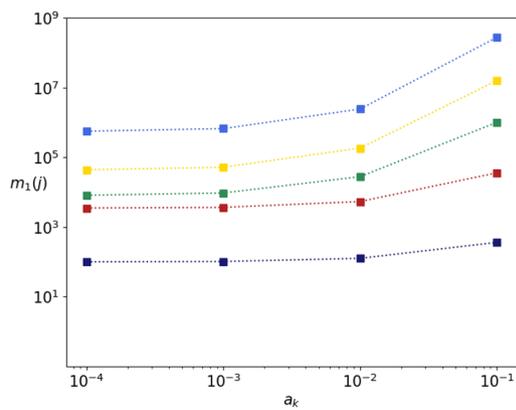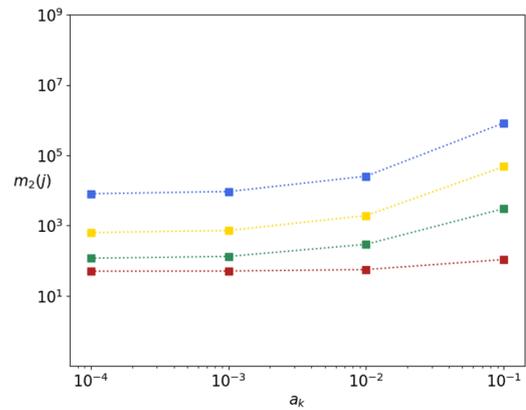
(a) HSC1 progenitor for Symm1 model.
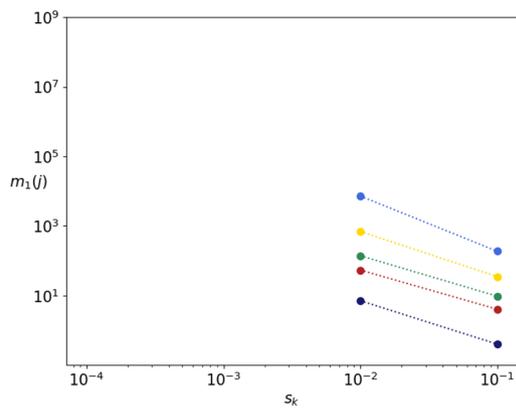
(b) HSC1 progenitor for SymmAll model.

(c) HSC2 progenitor for SymmAll model.

(d) HSC1 progenitor for AsymmAll model.

(e) HSC2 progenitor for AsymmAll model.

Figure 10. Mean number, $m_i(j)$, of cells in the progeny for an (**a**, **b**, **d**) HSC1 ($i = 1$) or (**c**, **e**) HSC2 ($i = 2$) progenitor, in compartments $j \in \{i, \ldots, 5\}$. In each scenario, we vary the corresponding rate (s_1 in Symm1, s_k, $k \in \{1, \ldots, N\}$ in SymmAll, and a_k, $k \in \{1, \ldots, N\}$ in AsymmAll), where only values leading to finite $m_i(j)$ are considered.
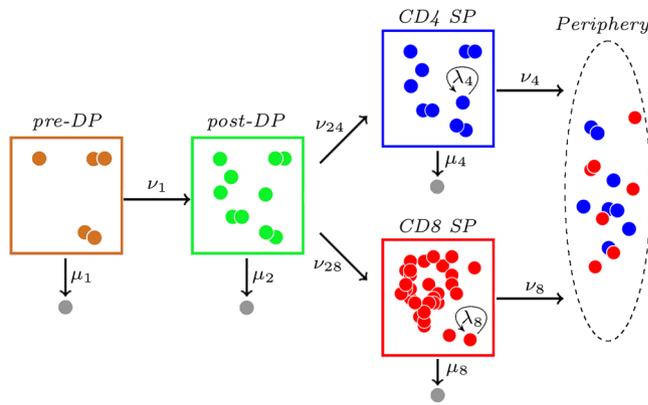
Figure 11. Thymic development model proposed in Ref.[14]. Grey cells represent death events.

$$\beta_1(1) = \frac{\mu_1}{\mu_1 + \nu_1},$$

$$\beta_1(2) = \frac{\nu_1}{\mu_1 + \nu_1} \frac{\mu_2}{\mu_2 + \nu_{24} + \nu_{28}},$$

$$\beta_1(4) = \frac{\nu_1}{\mu_1 + \nu_1} \frac{\nu_{24}}{\mu_2 + \nu_{24} + \nu_{28}} \frac{\mu_4}{\mu_4 + \nu_4},$$

$$\beta_1(8) = \frac{\nu_1}{\mu_1 + \nu_1} \frac{\nu_{28}}{\mu_2 + \nu_{24} + \nu_{28}} \frac{\mu_8}{\mu_8 + \nu_8},$$

$$\beta_1(4P) = \frac{\nu_1}{\mu_1 + \nu_1} \frac{\nu_{24}}{\mu_2 + \nu_{24} + \nu_{28}} \frac{\nu_4}{\mu_4 + \nu_4},$$

$$\beta_1(8P) = \frac{\nu_1}{\mu_1 + \nu_1} \frac{\nu_{28}}{\mu_2 + \nu_{24} + \nu_{28}} \frac{\nu_8}{\mu_8 + \nu_8}.$$

These analytical expressions allow us to perform a local sensitivity analysis by computing partial derivatives with respect to model parameters. For example, we have

$$\frac{\partial \beta_1(8P)}{\partial \nu_{24}} = \frac{\partial \beta_1(8P)}{\partial \mu_2} = -\frac{\nu_1 \nu_{28} \mu_8}{(\mu_1 + \nu_1)(\mu_2 + \nu_{24} + \nu_{28})^2(\mu_8 + \nu_8)}.$$

The proliferative potential of thymocytes during thymic development directly depends on them reaching the CD4 SP or CD8 SP compartment, where they are able to divide, before they exit to the periphery. Thus, the average number of divisions initiated by a single pre-DP thymocyte during its thymic development journey, $\eta_1 = \eta_1(4) + \eta_1(8)$, is given by

$$\eta_1(4) = \frac{\nu_1 \nu_{24}}{(\mu_1 + \nu_1)(\mu_2 + \nu_{24} + \nu_{28})} \frac{\lambda_4}{\mu_4 + \nu_4},$$

$$\eta_1(8) = \frac{\nu_1 \nu_{28}}{(\mu_1 + \nu_1)(\mu_2 + \nu_{24} + \nu_{28})} \frac{\lambda_8}{\mu_8 + \nu_8}.$$

Finally, the average lifespan of a pre-DP cell during thymic development (i.e., the mean time until it dies or it reaches the periphery) is given by

$$\tau_1 = \frac{1}{\mu_1 + \nu_1} \left[\frac{\nu_1}{\mu_2 + \nu_{24} + \nu_{28}} \left(\frac{\nu_{24}}{\mu_4 + \nu_4} + \frac{\nu_{28}}{\mu_8 + \nu_8} + 1 \right) + 1 \right].$$

We consider parameter values in Table 2 selected from Ref.[[14], Section 3.2], to compute the average lifespan of a pre-DP cell during its thymic development journey (until it dies or reaches the periphery), which corresponds to $\tau_1 = 2.84$ days. During its lifetime, a cell may undergo differentiation and proliferation, before dying in one of the compartments without ever reaching the periphery, or reaching the periphery either as a CD4 or CD8 cell. We show the predicted cell fate probabilities in Fig. 12. The most likely outcome corresponds to cell death,

Rate	μ_1	ν_1	μ_2	ν_{24}	ν_{28}	λ_4	λ_8	μ_4	μ_8	ν_4	ν_8
Value	0.263	0.137	1.369	0.07	0.054	0.216	0.093	0.04	0.11	0.21	0.14

Table 2. Parameter values from Ref.[[14], Section 3.2], in units $days^{-1}$.
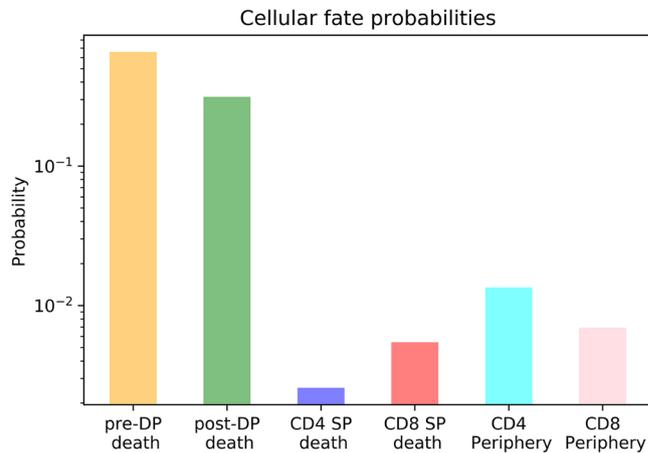
Figure 12. Probabilities of a single pre-DP cell to die in each of the compartments (pre-DP, post-DP, CD4 SP, or CD8 SP) before reaching the periphery, or to reach the periphery as a CD4 or CD8 SP cell. In particular, $\beta_1(1) = 0.6575$, $\beta_1(2) = 0.3140$, $\beta_1(4) = 0.0026$, $\beta_1(8) = 0.0055$, $\beta_1(4P) = 0.0135$ and $\beta_1(8P) = 0.0069$.

especially during the early stages (pre-DP and post-DP compartments). This agrees with existing evidence that most of negative selection occurs during the DP stages of development[55]. Once a cell reaches the CD4 SP compartment, it is more likely to reach the periphery than to die in that compartment, while these probabilities are comparable in the CD8 SP case.

In Table 3, we present the elasticities (i.e., normalised derivatives) of the probabilities $\beta_1(j)$, $j \in \{1, 2, 4, 8, 4P, 8P\}$, with respect to model parameters. This can be of particular relevance when parameters have been experimentally estimated with some uncertainty, so that one can assess the impact of perturbations in these values on specific model outputs. As expected, the division rates, λ_4, λ_8, do not affect the probability of death of the tracked cell starting as pre-DP. Moreover, the death rates $\mu_1, \mu_2, \mu_4, \mu_8$ positively contribute to the probability of the cell dying in the corresponding compartment $j \in \{1, 2, 4, 8\}$, while negatively contributing to the probability of the cell dying in other compartments. It is also worth noting that, for $j \in \{2, 4, 8\}$, the probability, $\beta_1(j)$, of the cell dying in that compartment is mainly affected by the differentiation rate into compartment j; that is, $\nu_1, \nu_{24}, \nu_{28}$, respectively. This can be understood since we are tracking a cell starting in compartment $i = 1$, and following its developmental journey across the sequence of compartments. Thus, the probability of dying in a compartment $j \neq 1$ is mainly determined by the differentiation rates of previous compartments.

The average number of division events performed by a single pre-DP cell is $\eta_1 = \eta_1(4) + \eta_1(8) = 0.0139 + 0.0046 = 0.0185$. This implies that out of 10^2 pre-DP cells starting the thymic development journey only (about) 2 cells are expected to be produced by cell division from the original cells, when visiting the CD4 SP or CD8 SP compartments. These small values are directly related to the small probabilities of reaching these compartments at all, so that the cell can actually divide. Our results here are in agreement with the results from Ref.[14], where the authors make use of a deterministic model to conclude that thymic development is a rather stringent process characterised by an extremely low success rate.

Discussion

We have presented a general model to characterise the stochastic journeys of cell progenies through compartments. Cells can divide, die or exit to adjacent compartments. We have derived analytical expressions for the mean number of cells in each compartment as a function of time, under different scenarios of interest (e.g., irreversible model, where differentiation cannot be reversed) and studied the progeny of a single progenitor cell

$\frac{\partial \beta_i(j)}{\partial \theta} / \frac{\beta_i(j)}{\theta}$	μ_1	ν_1	μ_2	ν_{24}	ν_{28}	λ_4	λ_8	μ_4	μ_8	ν_4	ν_8
$\beta_1(1)$	0.34	−0.34	0	0	0	0	0	0	0	0	0
$\beta_1(2)$	−0.66	0.66	0.08	−0.04	−0.04	0	0	0	0	0	0
$\beta_1(4)$	−0.66	0.66	−0.92	0.96	−0.04	0	0	0.84	0	−0.84	0
$\beta_1(8)$	−0.66	0.66	−0.92	−0.04	0.96	0	0	0	0.56	0	−0.56
$\beta_1(4P)$	−0.66	0.66	−0.92	0.96	−0.04	0	0	−0.16	0	0.16	0
$\beta_1(8P)$	−0.66	0.66	−0.92	−0.04	0.96	0	0	0	−0.44	0	0.44

Table 3. Elasticities for the probabilities $\beta_1(j)$, $j \in \{1, 2, 3, 8, 4P, 8P\}$, with respect to parameter $\theta \in \{\mu_1, \nu_1, \mu_2, \nu_{24}, \nu_{28}, \lambda_4, \lambda_8, \mu_4, \mu_8, \nu_4, \nu_8\}$. They are given by $(\partial \beta_1(j)/\partial \theta)/(\beta_i(j)/\theta)$, with cell fate probabilities (in rows) and model parameters (in columns).

in terms of the probability generating function and summary statistics appropriately defined. The analysis allows us to track the journey of a lifeline across the system of compartments. We have then calculated its lifetime, its proliferative potential and the probability of different cell fates. We have used case studies to illustrate the applicability of our techniques and the impact of model parameters on the corresponding summary statistics.

The analysis carried out in Section "Lifeline analysis" sheds light on the dynamics of the cell population by analysing the journey followed by a single cell. Moreover, this technique is rather "flexible" since the first-step arguments in which it relies can be easily extended to other compartmental topologies, as we have briefly illustrated in the third case study (see Section "Tracking a thymocyte during its development"). Moreover, novel labelling and barcoding techniques provide an increasing amount of data[56–59], which could be compared to this type of model predictions.

Our results also highlight the significant role that symmetric and asymmetric division events can play in these systems, when compared to self-renewal. In particular, we have shown in Section "Asymmetric and symmetric division: the case of four compartments" how symmetric division events can significantly affect the dynamics of the system, potentially moving it from unbounded growth to extinction. Increasing the asymmetric division rates does not change the late time behaviour. Still, it can delay population extinction by increasing the number of cells arising over time across compartments. On the other hand, for a fixed division rate, ω, and different probabilities of each type of division event (self-renewal, p_{SR}; symmetric division, p_{SD}; asymmetric division, p_{AD}), compartmental systems where symmetric or asymmetric division is more likely lead to smaller cell populations and faster dynamics to extinction or steady-state, compared to systems where self-renewal is the dominant division process. Interestingly, increasing the probability of symmetric or asymmetric division leads to smaller progenies from a single progenitor cell, while maximising the size of the fully differentiated (or terminal) population. We note that, in Section "Asymmetric and symmetric division: the case of four compartments" we set all differentiation rates to be the same across compartments, since the focus was on studying the impact of symmetric/asymmetric division on the cell population dynamics. However, it is to be expected that significant heterogeneity in differentiation rates across compartments could make specific compartments more/ less important in the dynamics. In particular, compartments with significantly large differentiation rates would lead to shorter residence times, which would likely imply that other events which can occur in these compartments (e.g., division) would become less relevant.

A particular limitation of our approach is that it relies on cells behaving independently from each other, as in the theory of branching processes. While this may be valid in some situations, for example when studying small cell populations where tissue growth control is only through feedback by the cell density, it might not be valid in other scenarios (e.g., for large cell populations, or under competition for resources, where logistic growth-type models might be more appropriate). This, in turn, is related to the fact that the corresponding ODE system for the average number of cells in each compartment is linear. Cell independence (or linearity) allows one to implement techniques from the theory of branching processes, and makes the single-cell analysis proposed here feasible, since we identify lifelines amongst the dynamics of all the cells in the compartmental system. Relaxing this particular assumption is, thus, the aim of future work.

Data availability
Computer codes to generate Figures 5–10 and Figure 12 can be accessed at https://github.com/matml/Journey_Of_A_Cell_Across_A_Sequence_Of_Compartments. There is no other relevant data related to this study.

References
1. Martin, G. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci.* **78**(12), 7634–7638 (1981).
2. Evans, M. & Kaufman, M. Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292**(5819), 154–156 (1981).
3. Zhang, H. *et al.* Modelling epidermis homoeostasis and psoriasis pathogenesis. *J. R. Soc. Interface* **12**, 20141071 (2015).
4. Fuchs, E. Skin stem cells: Rising to the surface. *J. Cell Biol.* **180**, 273–284 (2008).
5. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
6. Murray, P. J. *et al.* Comparing a discrete and continuum model of the intestinal crypt. *Phys. Biol.* **8**, 026011 (2011).
7. Potten, C., Gandara, R., Mahida, Y., Loeffler, M. & Wright, N. The stem cells of small intestinal crypts: Where are they?. *Cell Prolif.* **42**, 731–750 (2009).
8. Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518**(7540), 542–546 (2015).
9. Barile, M. *et al.* Hematopoietic stem cells self-renew symmetrically or gradually proceed to differentiation. *Available at SSRN 3787896* (2020).
10. Dingli, D., Traulsen, A. & Pacheco, J. M. Compartmental architecture and dynamics of hematopoiesis. *PLoS ONE* **2**, e345 (2007).
11. Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).
12. Michor, F., Iwasa, Y., Lengauer, C. & Nowak, M. A. Dynamics of colorectal cancer. In *Seminars in Cancer Biology* Vol. 15 484–493 (Elsevier, 2005).
13. Yates, A. Theories and quantification of thymic selection. *Front. Immunol.* https://doi.org/10.3389/fimmu.2014.00013 *(2014).*
14. Sawicka, M. *et al.* From pre-DP, post-DP, SP4, and SP8 thymocyte cell counts to a dynamical model of cortical and medullary selection. *Front. Immunol.* https://doi.org/10.3389/fimmu.2014.00019 *(2014).*
15. Pham, K. *et al.* Asymmetric cell division during T cell development controls downstream fate. *J Cell Biol.* **210**(6), 933–950. https://doi.org/10.1083/jcb.201502053 (2015).
16. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
17. Paley, M. A. *et al.* Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science* **338**, 1220–1225. https://doi.org/10.1126/science.1229620 (2012).
18. Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
19. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).

20. McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu. Rev. Immunol.* **37**, 457–495 (2019).
21. Montagna, W., Kligman, A. & Carlisle, K. Overview of skin. In *Atlas of Normal Human Skin* 3–5 (Springer, 1992).
22. Marjanovic, N. D., Weinberg, R. A. & Chaffer, C. L. Cell plasticity and heterogeneity in cancer. *Clin. Chem.* **59**, 168–179 (2013).
23. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**(8), 486–499. https://doi.org/10.1038/nri3862 (2015).
24. Klein, A. & Simons, B. Universal patterns of stem cell fate in cycling adult tissues. *Development* https://doi.org/10.1242/dev.060103 (2011).
25. Werner, B., Dingli, D., Lenaerts, T., Pacheco, J. M. & Traulsen, A. Dynamics of mutant cells in hierarchical organized tissues. *PLoS Comput. Biol.* **7**, e1002290 (2011).
26. Hodzic, E. Single-cell analysis: Advances and future perspectives. *Bosn. J. Basic Med. Sci.* **16**(4), 313 (2016).
27. Perié, L. *et al.* Determining lineage pathways from cellular barcoding experiments. *Cell Rep.* **6**(4), 617–624 (2014).
28. Werner, B., Dingli, D., Lenaerts, T., Pacheco, J. M. & Traulsen, A. Dynamics of mutant cells in hierarchical organized tissues. *PLoS Comput. Biol.* **7**, e1002290 (2011).
29. Werner, B., Dingli, D. & Traulsen, A. A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues. *J. R. Soc. Interface* **10**, 20130349 (2013).
30. Werner, B. *et al.* The cancer stem cell fraction in hierarchically organized tumors can be estimated using mathematical modeling and patient-specific treatment trajectories. *Cancer Res.* **76**, 1705–1713 (2016).
31. Johnston, M. D., Maini, P. K., Chapman, S. J., Edwards, C. M. & Bodmer, W. F. On the proportion of cancer stem cells in a tumour. *J. Theor. Biol.* **266**, 708–711 (2010).
32. Kaveh, K., Kohandel, M. & Sivaloganathan, S. Replicator dynamics of cancer stem cell: Selection in the presence of differentiation and plasticity. *Math. Biosci.* **272**, 64–75 (2016).
33. Kaveh, K. Stem cell evolutionary dynamics of differentiation and plasticity. *Curr. Stem Cell Rep.* **3**, 366–372 (2017).
34. Weekes, S. L. *et al.* A multicompartment mathematical model of cancer stem cell-driven tumor growth dynamics. *Bull. Math. Biol.* **76**, 1762–1782 (2014).
35. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446**, 185–189 (2007).
36. Dingli, D., Traulsen, A. & Michor, F. (A) symmetric stem cell replication and cancer. *PLoS Comput. Biol.* **3**, e53 (2007).
37. Mamis, K., Zhang, R. & Bozic, I. Stochastic model for cell population dynamics quantifies homeostasis in colonic crypts and its disruption in early tumorigenesis. *Proc. R. Soc. B* **290**, 20231020 (2023).
38. Shahriyari, L. & Komarova, N. L. Symmetric vs. asymmetric stem cell divisions: An adaptation against cancer?. *PloS One* **8**, e76195 (2013).
39. Deshmukh, S. & Saini, S. Phenotypic heterogeneity in tumor progression, and its possible role in the onset of cancer. *Front. Genet.* **11**, 604528 (2020).
40. Kimmel, M. & Axelrod, D. E. *Branching Processes in Biology* 9th edn. (Springer, 2002).
41. Watson, H. W. & Galton, F. On the probability of the extinction of families. *J. Anthropol. Inst. G. B. Irel.* **4**, 138–144 (1875).
42. de la Higuera, L. *et al.* Fate of a naive T cell: A stochastic journey. *Front. Immunol.* **10**, 194 (2019).
43. Nordon, R. E., Ko, K. H., Odell, R. & Schroeder, T. Multi-type branching models to describe cell differentiation programs. *J. Theor. Biol.* **277**(1), 7–18. https://doi.org/10.1016/j.jtbi.2011.02.006 (2011).
44. Matis, J. H. & Hartley, H. O. *Stochastic Compartmental Analysis: Model and Least Squares Estimation from Time Series Data* (International Biometric Society, 1971).
45. Teschl, G. *Ordinary Differential Equations and Dynamical Systems* Vol. 140 (American Mathematical Soc., 2012).
46. Dreher, R. Modified Bateman solution for identical eigenvalues. *Ann. Nucl. Energy* **53**, 427–438. https://doi.org/10.1016/j.anucene.2012.06.019 (2013).
47. Conte, S. & De Boor, C. *Elementary Numerical Analysis: An Algorithmic Approach* (McGraw-Hill, 1972).
48. Thomas, L. H. *Elliptic Problems in Linear Difference Equations Over a Network* 71 (Watson Sci. Comput. Lab. Rept., Columbia University, 1949).
49. Singmaster, D. An elementary evaluation of the Catalan numbers. *Am. Math. Mon.* **85**, 366–368 (1978).
50. Castro, M., López-García, M., Lythe, G. & Molina-París, C. First passage events in biological systems with non-exponential inter-event times. *Sci. Rep.* **8**, 1–16 (2018).
51. Abate, J. & Whitt, W. A unified framework for numerically inverting Laplace transforms. *INFORMS J. Comput.* **18**, 408–421 (2006).
52. Seita, J. & Weissman, I. L. Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* https://doi.org/10.1002/wsbm.86 (2011).
53. Wilson, A., Laurenti, E. & Trumpp, A. Balancing dormant and self-renewing hematopoietic stem cells. *Curr. Opin. Genet. Dev.* https://doi.org/10.1016/j.gde.2009.08.005 (2009).
54. Bowman, R. L., Busque, L. & Levine, R. L. Clonal hematopoiesis and evolution to hematopoietic malignancies. *Cell Stem Cell* https://doi.org/10.1016/j.stem.2018.01.011 (2018).
55. Yates, A. Theories and quantification of thymic selection. *Front. Immunol.* **5**, 13 (2014).
56. Johnson, J. *et al.* Quantitative analysis of complex glioma cell migration on electrospun polycaprolactone using time-lapse microscopy. *Tissue Eng. Part C Methods* **15**, 531–540 (2009).
57. Westera, L. *et al.* Closing the gap between T-cell life span estimates from stable isotope-labeling studies in mice and humans. *Blood J. Am. Soc. Hematol.* **122**, 2205–2212 (2013).
58. Knabel, M. *et al.* Reversible MHC multimer staining for functional isolation of T-cell populations and effective adoptive transfer. *Nat. Med.* **8**, 631–637 (2002).
59. Krutzik, P. O. & Nolan, G. P. Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. *Nat. Methods* **3**, 361–368 (2006).

Acknowledgements

Author contributions

All authors conceived the mathematical models. H.D., F.F. and M.L.G. carried out the analysis, obtained the numerical results and drafted a first version of the manuscript. All authors contributed to and reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-63500-2.

Correspondence and requests for materials should be addressed to M.L.-G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.