Mathematical modelling of the hematopoietic stem cell-niche system

Clonal dominance based on stem cell fitness

Pedersen, Rasmus Kristoffer; Andersen, Morten; Stiehl, Thomas; Ottesen, Johnny T.

Published in:
Journal of Theoretical Biology

DOI:
10.1016/j.jtbi.2021.110620

Publication date:
2021

Document Version
Peer reviewed version

Citation for published version (APA):
Pedersen, R. K., Andersen, M., Stiehl, T., & Ottesen, J. T. (2021). Mathematical modelling of the hematopoietic stem cell-niche system: Clonal dominance based on stem cell fitness. *Journal of Theoretical Biology*, *518*(518), Article 110620. https://doi.org/10.1016/j.jtbi.2021.110620

Mathematical modelling of the hematopoietic stem cell-niche system: Clonal dominance based on stem cell fitness.

Rasmus Kristoffer Pedersen[1,3], Morten Andersen[1], Thomas Stiehl[2], and Johnny T. Ottesen[1]

[1]Roskilde University, Denmark
[2]Heidelberg University, Germany
[3]Region Sjælland, Denmark

Abstract

Human blood cell production is maintained by hematopoietic stem cells (HSC) which give rise to all types of mature blood cells. Experimental observation of HSC in their physiologic bone-marrow microenvironment, the so-called stem cell niche, is challenging. Therefore, the details of HSC dynamics and the cellular interactions in the stem cell niche remain elusive. Mutations that give a competitive advantage to the mutated cells are the cause of clinical challenges when treating HSC-derived malignancies such as acute myeloid leukemia (AML) or the myeloproliferative neoplasms (MPNs). To investigate the significance of the interaction between the HSC and the stem cell niche in these malignancies, we propose and analyse a mechanism-based mathematical model of HSC dynamics within the bone-marrow microenvironment.

The model is based on the central hypothesis that HSC self-renewal depends on the niche. In the model, the interaction of HSC with specific niches located in the bone marrow are key to the indefinite renewal necessary for long-term maintenance of blood-production. We formulate a general model of n distinct clones that differ with respect to cell properties. We identify an attractive trapping region and compute and classify all steady states. A concept of HSC fitness naturally arises from the model analysis. HSC fitness is found to determine the asymptotic behaviour of the model, as the only steady state that is locally stable is related to the HSC-clone with the highest fitness.

Based on biological assumptions about HSC, we propose two reduced models of different complexity. A thorough mathematical analysis reveals that both reduced models have the same asymptotic behaviour as the full model.

We compare the simpler of the two models, a logistic equation of the disease burden, to clinical data of MPN-patients. The reduced model is found to agree well with data and suggests a simple interpretation and possible prediction of patient prognosis.

The proposed mathematical model and the reduced forms have the potential to provide insights into the regulation of HSC dynamics and blood cell formation, and ultimately for future advances in treatment of hematologic malignancies.

1 Introduction

Human blood consists of a vast amount of blood cells. As these mature blood cells (such as e.g. red blood cells, thrombocytes or leukocytes) die by programmed cell death or other mechanisms, a large number of new cells must be produced each day. The process of blood cell production, referred to as hematopoiesis, originates in the pool of hematopoietic stem cells (HSC). Through many steps of cell-division, self-renewal and differentiation, the diverse types of blood cells are produced, without exhaustion of the HSC population. While much research has been done in an effort to uncover the cell intrinsic regulatory mechanisms of HSC, open questions about their properties and functions still remain.

One particular area where an understanding of HSC-dynamics has a significant impact is the research of hematopoietic malignancies, such as acute myeloid leukemia (AML), chronic myeloid leukemia (CML) or the myeloproliferative neoplasms (MPNs). These diseases are believed to arise from malignant stem cells (Reya et al., 2001) that give rise to malignant cells which trigger disease-specific symptoms, such as e.g. an excess of thrombocytes. Understanding how malignant stem cells differ from wild-type HSC is crucial for efficient treatment (Dingli and Michor, 2006) and hence patient prognosis. HSC-specific niches have been in the focus of some research as they have been found to induce cellular quiescence, which can explain why some malignant stem cells respond poorly to treatment (Ishikawa et al., 2007).

Somatic mutations in the HSC genome can result in a subpopulation of HSC, referred to as a HSC clone, that differ from the normal HSC. As mutations accumulate through ageing, clonal hematopoiesis in which a substantial part of the production of blood cells is due to a single clone is expected to arise in as many as 10% of the population of people over 65 years of age. (Genovese et al., 2014). While clonal hematopoiesis by itself does not necessarily lead to symptoms of disease, it gives rise to the clinical concept of clonal hematopoiesis of indeterminate potential (CHIP) which can be a precursor or early stage of hematopoietic malignancy (Jaiswal and Ebert, 2019). Determining the significance of mutations and CHIP can be highly relevant in regards to the clinical question of understanding the risk of hematopoietic malignancies, particularly in an ageing population.

Despite years of comprehensive research on the behaviour and dynamics of HSC, both mathematically and biologically, a complete quantitative description still remains unclear. This leaves some open questions about HSC yet to be answered, such as:

- Which role do the HSC-specific niches play in maintaining HSC and blood cell counts?
- Which factors determine the outcome of competition between multiple distinct clones within the bone marrow microenvironment?
- How do mutational changes to HSC properties lead to the wide range of disease dynamics observed in patients with hematologic malignancies?
- Why are some clinical interventions highly successful for specific HSC-derived malignancies, yet less successful for other similar malignancies?
- Can insights about one malignancy guide the clinical management of other, potentially similar, malignancies?

In an effort to explore these questions, as well as many others, mathematical models have been developed throughout the literature. These include models of ordinary differential equations (Stiehl et al., 2015; Wang et al., 2017; Stiehl et al., 2020; Dingli and Michor, 2006; Andersen et al., 2017; Komarova and Wodarz, 2007; Ashcroft et al., 2017), delay differential equations (Colijn and Mackey, 2005; Park et al., 2019) as well as stochastic models (Roeder and Loeffler, 2002; Catlin et al., 2011). We refrain from giving a complete review of such work, but direct the reader to the work of Clapp and Levy (2015) or to a more general review on cancer modelling due to Altrock et al. (2015).

The mathematical models of hematologic malignancies and HSC behaviour found in the literature differ greatly in complexity, and hence the applicability of the models also varies. While complex models provide a detailed picture of a system, simpler models allow more directly for comparison with clinical data. Hence a dilemma arises where a model must have enough detail to be biologically accurate, but still simple enough to be useful in a clinical setting. Constructing complex models based on biological ideas and hypotheses allows for investigations about the nature of those ideas and hypotheses. By reducing and simplifying a complex model in mathematically or biologically reasonable ways, a simpler model arises. Such simpler models allow for comparison with data which would have been inconclusive when based on the complex model. In this way simplification helps to identify the data necessary to distinguish between different complex models. Hence, model reduction, simplification and subsequent comparison with experimental data both clarify the applicability of the model and pave the way for future experiments that can determine the validity of the hypotheses that the model was based on.

In this paper we present a model of HSC-behaviour, which captures the most significant parts of the biology, as well as the dynamical behaviour of blood cell production observed in both healthy and diseased individuals. The model is based on the dynamics of HSC within the bone-marrow microenvironment and builds on a central hypothesis about limited HSC renewal following detachment from HSC-specific niches. The processes impairing self-renewal when HSC are removed from their microenvironment are not well understood. Different from previous models of HSC dynamics (Stiehl et al., 2020; Ashcroft et al., 2017; Becker et al., 2019), we here propose that HSC can perform a limited number of divisions after detaching from the niche and then either differentiate or reattach to the niche. Reattachment to the niche is required to maintain stemness and to become reactivated at a later time.

A simplified version of the model has previously been presented by the authors (Pedersen et al., nd), and was used to investigate medical interventions such as HSC mobilization and transplantation. In particular, we found evidence that when suggesting novel treatment strategies of HSC-derived malignancies, careful attention must be paid to the treatment-induced changes in healthy and malignant stem cell dynamics to avoid potential harm.

Through mathematical analysis of the model, we determine the dynamics for non-negative initial conditions of a general model of n distinct HSC clones. By assuming that most HSC remain quiescent or inactive for extended periods of time, we propose a reduced model. The reduced form maintains the same general structure and dynamics as the full model, but allows for a more direct comparison with experimental data from literature.

Following still further simplifications, the simplest form of the model is compared to clinical data. In the most reduced form, the model is in agreement with data, suggesting that the proposed model is useful, albeit overparametrized in the full form with the current data available.

2 Model presentation

We propose a model of hematopoetic stem cells (HSC) and in particular their interaction with the bone marrow. Such a model has previously been developed by us (Pedersen et al., nd), and the present work is a generalization of that model. The biological foundation and assumptions are described here:

- HSC-specific niches are located in the bone marrow microenvironment. The specific cellular composition of the niches is not considered, and rather we consider an abstract notion of niches, assumed to be well-mixed throughout the bone marrow. It has been suggested that niche cells dynamically respond to the presence of HSC, and fluctuate in numbers (Becker et al., 2019). We hypothesize that the effect is minor and assume constant counts of niche-cells.

- HSC bind to the niches and remain quiescent for extended periods of time. The nature of how quiescence of HSC is maintained by the niches is a complex picture involving many different types of signalling (Nie et al., 2008; Nilsson et al., 2005), and is assumed to be necessary for maintaining stem cell function (Kumar and Geiger, 2017; Vaidya and Kale, 2015; Zhang and Gao, 2016). We denote niche-bound HSC as N, and denote the remaining free niches N_E.

- Niche-bound cells are assumed to detach from the niches with rate u, either spontaneously or through activation. The specifics of how this activation occurs are assumed not to impact on the behaviour of the resulting free HSC. We denote the activated free HSC as A. We assume that activated HSC can reattach to the empty niches N_E with rate b_A. These two processes can be written as:

$$N \xrightarrow{u} A + N_E \tag{1}$$

$$N_E + A \xrightarrow{b_A} N \tag{2}$$

- Upon detaching from the niche, the free HSC lose their quiescence and enter the cell-cycle (Kumar and Geiger, 2017). The detached cells are primed to undergo either self-renewal or differentiation, possibly due to signalling from the niche prior to release (Wilson and Trumpp, 2006). We assume that when HSC differentiate into more mature cells, they cannot reacquire stemness. We focus on the remaining stem cells, and differentiation is simply considered a loss of cells. If differentiation occurs with rate d_A this can be written as:

$$A \xrightarrow{d_A} \emptyset \tag{3}$$

  Note that this may include death of the HSC. However, under normal circumstances HSC death is assumed to occur for a negligible fraction of HSC.

  If a free HSC self-renews, it divides into two HSC. Assuming the rate of division-related mutations is negligible, the two resulting cells are identical. We hypothesize that the resulting HSC are limited in their ability to further self-renew if they do not reattach to the niche. HSC following division that have to reattach to the niche to restore their full self-renewing potential are referred to as inhibited or exhausted HSC. For the the most general formulation, we consider one free HSC to self-renew with rate r and that the process results in 2γ inhibited HSC where $\gamma \geq 1$. In the supplementary material A we argue for the appropriateness of including the factor $\gamma \in \mathbb{R}$. The process can be written as:

$$A \xrightarrow{r} 2\gamma I \tag{4}$$

  where I denotes the inhibited HSC.

- The inhibited HSC cannot self-renew, but are still assumed to differentiate into more mature cells such as progenitors. This occurs with rate d_I, and as above we do not keep track of the resulting cells:

$$I \xrightarrow{d_I} \emptyset \tag{5}$$

  Inhibited HSC are hypothesized to attach to free niches, thereby gaining the ability to self-renew if they detach again. The attachment occurs with rate b_I and thus we write the process as:

$$I + N_E \xrightarrow{b_I} N \tag{6}$$

The above processes are modelled as a system of differential equations, describing the rate of change of free niches, N_E, and the different states of HSC, N, A and I.

In the bone-marrow microenvironment, HSC with different cell-specific properties can co-exist for extended periods of time. We refer to HSC subpopulations with persistent and inheritable differences in cell properties as different clones. To define our model in the most general form, we model n different clones. The rates involved in the processes described above can all differ from one clone to the other, and hence we write each clone-specific rate with a subscript. Note that different clones interact only through the niches. No direct interaction between cells of different clones are considered in the model. A schematic representation of the model for $n = 2$ is shown in figure 1. The complete n-clone system of differential equations has $3n + 1$ dimensions, and is given by:

$$\dot{N_E} = \sum_{i=1}^{n} u_i N_i - N_E \sum_{i=1}^{n} \left(b_{I_i} I_i + b_{A_i} A_i\right) \tag{7a}$$

$$\dot{N_j} = b_{I_j} N_E I_j + b_{A_j} N_E A_j - u_j N_j \tag{7b}$$

$$\dot{I_j} = 2\gamma r_j A_j - b_{I_j} N_E I_j - d_{I_j} I_j \tag{7c}$$

$$\dot{A_j} = u_j N_j - b_{A_j} N_E A_j - r_j A_j - d_{A_j} A_j \tag{7d}$$

where j refers to the j^{th} clone and $\dot{}$ represents the derivative with respect to time t, $\frac{d}{dt}$. N_E denotes the abundance of empty niches, N_j are niches occupied by HSC of the j^{th} clone while A_j and I_j are free HSC of the j^{th} clone, in the active and inactive state, respectively. Based on their biological meaning, all parameters are non-negative: $u_j, b_{I_j}, b_{A_j}, r_j, d_{I_j}, d_{A_j} \geq 0$ for all j. A unique solution to system (7) exists and the solutions $N_E(t), N_j(t), I_j(t), A_j(t)$ are C^1, since the right-hand sides of system (7) are continuous and fulfil a Lipschitz condition in the variables. As negative cell-counts are biologically meaningless, we define *feasibility* of solutions.

Definition 2.1: Feasibility

A set of solutions N_E, N_j, I_j, A_j is *feasible* if all variables N_E, N_j, I_j, A_j are non-negative for all j.

For non-negativity (and hence feasibility) of solutions, the time-derivatives for each variable must be non-negative when the given variable is zero. For all variables non-negative we find:

$$\dot{N}_{E|N_E=0} = \sum_{i=1}^{n} u_i N_i \geq 0 \tag{8a}$$

$$\dot{N}_{j|N_j=0} = N_E \left(b_{A_j} A_j + b_{I_j} I_j\right) \geq 0 \tag{8b}$$

$$\dot{I}_{j|I_j=0} = 2\gamma r_j A_j \geq 0 \tag{8c}$$

$$\dot{A}_{j|A_j=0} = u_j N_j \geq 0 \tag{8d}$$

for all j. Thus, due to the existence and uniqueness of solutions described above, the solutions of the system will remain non-negative for all $t > 0$, given non-negative initial conditions at $t = 0$.

From equations (7a) and (7b), $\dot{N_E} + \sum_{i=1}^{n} \dot{N_i} = 0$, and we define $K = N_E + \sum_{i=1}^{n} N_i$. As all variables are non-negative, $K \geq 0$, and in general we assume $K > 1$. Eliminating N_E, the system of equations reduces to a $3n$ dimensional system:

$$\dot{N_j} = b_{I_j} \left(K - \sum_{i=1}^{n} N_i\right) I_j + b_{A_j} \left(K - \sum_{i=1}^{n} N_i\right) A_j - u_j N_j \tag{9a}$$

$$\dot{I_j} = 2\gamma r_j A_j - b_{I_j} \left(K - \sum_{i=1}^{n} N_i\right) I_j - d_{I_j} I_j \tag{9b}$$

$$\dot{A_j} = u_j N_j - b_{A_j} \left(K - \sum_{i=1}^{n} N_i\right) A_j - r_j A_j - d_{A_j} A_j \tag{9c}$$

for all j. In addition, $N_E = K - \sum_{i=1}^{n} N_i$. Assuming feasible initial conditions, $N_E \geq 0$ restricts the range of $N_j(t)$ to $[0, K]$ for all j. In fact, $N_E \geq 0$ implies $\sum_{i=1}^{n} N_i(t) \leq K$. For feasible initial conditions, the range of $I_j(t)$ and $A_j(t)$ is $[0, \infty)$ for all j. This reformulation is an equivalent system, so results and conclusions based on one of the two formulations also holds for the other formulation. For this reason, parts of the analysis use the formulation in equations (7), while other parts make use of equations (9) to ease analysis.

Default parameters, determined in our previous work (Pedersen et al., nd), are given in table 1.

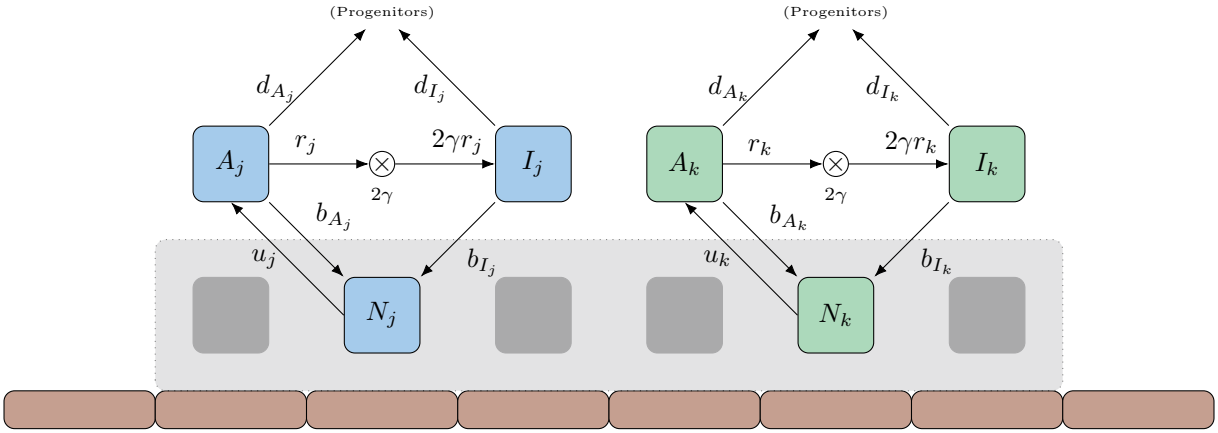


Figure 1: **Compartment diagram of the model.** The diagram depicts two distinct clones and the (indirect) competition through the niche. The two clones are denoted by subscript j and k respectively. Each box illustrates a variable while arrows illustrate the relationship between variables. The circle with the cross, $\times$, signifies a multiplication with 2γ. Note that I and A cannot change to N when there are no free niches, illustrated by the darkened boxes. The red rectangles at the bottom illustrate the bone-marrow cells that are part of the HSC niches.

K	15000 cells	u_j	0.04 day^{-1}
r_j	2.32 day^{-1}	b_{I_j}	0.96 day^{-1}
d_{A_j}	2.06 day^{-1}	d_{I_j}	3.77 day^{-1}

Table 1: **Default parameters for the proposed model.** Parameter-values were determined in our previous work (Pedersen et al., nd). Unless otherwise noted, parameters are equal for all clones j. While $b_{A_j} = 0$ was previously assumed, we here set $b_{A_j} = 0.01$ day^{-1}.

3 Model analysis

3.1 Identifying an attractive trapping region

We showed above that the solutions of equations (9) remain non-negative for non-negative initial condition, and that $N_j(t)$ is restricted to the domain $[0, K]$ for all j. In this section we show that an attractive trapping region exists in which solutions with initial conditions inside the region stay within the region.

An upper limit for the sum of niche-bound cells is $\sum_{i=1}^{n} N_i = K$, and thus for any given clone $N_j = K - \sum_{k\neq j} N_k \geq 0$. Hence, N_j is bound from above, since due to equation (9a),

$$\dot{N}_j|_{\sum_{i=1}^{n} N_i = K} = -u_j N_j \leq 0 \tag{10}$$

Note that $\dot{N}_j = 0$ exactly when $N_j = 0$.

Extending this, note that:

$$\sum_{i=1}^{n} \dot{N}_i|_{\sum_{i=1}^{n} N_i = K} = -\sum_{i=1}^{n} u_i N_i < 0 \tag{11}$$

since not all $u_j N_j$ are zero when $\sum_{i=1}^{n} N_i = K$.

We define $\hat{A}_{j,\epsilon} = \frac{u_j K}{r_j + d_{A_j}} + \epsilon$ for any $\epsilon \geq 0$. Thus:

$$\dot{A}_j|_{A_j = \hat{A}_{j,\epsilon}} = -u_j(K - N_j) - (r_j + d_{A_j})\epsilon - b_{A_j}(K - \sum_{i=1}^{n} N_i)\left(\frac{u_j K}{r_j + d_{A_j}} + \epsilon\right) \tag{12}$$

$$\leq -(r_j + d_{A_j})\epsilon \leq 0 \tag{13}$$

since $\sum_{i=1}^{n} N_i \leq K$ and $N_j \leq K$. Hence $\dot{A}_j \leq 0$ for $A_j = \hat{A}_{j,0}$ and $\dot{A}_j < 0$ for $A_j \geq \hat{A}_{j,\epsilon}$ for any $\epsilon > 0$.

Consider an interval $\mathbb{A} = [\hat{A}_{j,\epsilon}, \hat{A}_{j,\epsilon} + \Delta]$ with $\Delta > 0$ on which $\dot{A}_j < 0$. Due to the extreme value theorem, there exists a finite μ such that $-\dot{A}_j \geq \mu, \forall A_j \in \mathbb{A}$, since $\mathbb{A}$ is a compact subset. In particular

$$\inf_{\mathbb{A}}(-\dot{A}_j) = \min_{\mathbb{A}}(-\dot{A}_j) = \mu > 0 \tag{14}$$

and hence $\frac{1}{-\dot{A}_j} \leq \frac{1}{\mu}$ holds for all $A_j \in \mathbb{A}$.

To determine the time t_A it takes for the flow from $A_j(0) = \hat{A}_{j,\epsilon} + \Delta$ to reach $A_j(t_A) = \hat{A}_{j,\epsilon}$ we integrate over the interval:

$$t_A = \int_0^{t_A} dt = \int_{A_j(0)}^{A_j(t_A)} \frac{1}{\dot{A}_j} dA_j = \int_{\hat{A}_{j,\epsilon}}^{\hat{A}_{j,\epsilon}+\Delta} \frac{1}{-\dot{A}_j} dA_j \leq \frac{\Delta}{\mu} \tag{15}$$

and hence t_A is finite. Thus for any $\Delta > 0$, any point with $A_j(0) = \hat{A}_{j,\epsilon} + \Delta$ will move along a trajectory that enters the region $A_j \leq \hat{A}_{j,\epsilon}$ in finite time for any $\epsilon > 0$.

For I_j, we define $\hat{I}_{j,\epsilon} = \frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}}\epsilon + \epsilon$ for any $\epsilon \geq 0$. Thus, for both $A_j \leq \hat{A}_{j,\epsilon}$ and $I_j \geq \hat{I}_{j,\epsilon}$:

$$\dot{I}_j|_{I_j = \hat{I}_{j,TR,\epsilon}} \leq 2\gamma r_j \left(\frac{u_j K}{r_j + d_{A_j}} + \epsilon\right) - \left(b_{j_I}\left(K - \sum_{i=1}^{n} N_i\right) + d_{I_j}\right)\left(\frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}}\epsilon + \epsilon\right) \tag{16}$$

$$\leq -d_{I_j}\epsilon \leq 0 \tag{17}$$

since $\sum_{i=1}^{n} N_i \leq K$. Hence for $A_j \leq \hat{A}_{j,\epsilon}$ and $I_j \geq \hat{I}_{j,\epsilon}$, $\dot{I}_j \leq 0$, with $\dot{I}_j < 0$ for all $\epsilon > 0$.

Through an argument similar to the one above, it can be shown that trajectories with $I_j(0) > \hat{I}_{j,\epsilon}$ will follow a trajectory that enters the region $I_j \leq \hat{I}_{j,\epsilon}$ in finite time for any $\epsilon > 0$.

Theorem 3.1: Trapping region TR_ϵ

Defining $\hat{A}_{j,\epsilon} = \frac{u_j K}{r_j + d_{A_j}} + \epsilon$ and $\hat{I}_{j,\epsilon} = \frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}}\epsilon + \epsilon$ for any $\epsilon \geq 0$, a trapping region for equations (9) is given by:

$$TR_\epsilon = \left\{ \prod_{j=1}^{n} (N_j, A_j, I_j) \in \mathbb{R}^{3n} : N_j \geq 0, A_j \geq 0, I_j \geq 0, \sum_{j=1}^{n} N_j \leq K, A_j \leq \hat{A}_{j,\epsilon}, I_j \leq \hat{I}_{j,\epsilon}, j = 1, \ldots, n \right\} \tag{18}$$

for any $\epsilon \geq 0$. For $\epsilon > 0$, TR_ϵ is an attracting trapping region.

3.2 Steady states

To understand the asymptotic behaviour of the trajectories of the model, we look for steady states in TR_0. All points in TR_0 have feasible values of the variables, and hence existence of such steady states also implies feasibility of the steady state. We denote the steady state values with an asterisk (*).

For brevity, we use N_E^* as the empty niches in a given steady state, i.e. $N_E^* = K - \sum_{i=1}^{n} N_i^*$. From equations (9) it is seen trivially that:

$$A_j^* = \frac{u_j N_j^*}{r_j + d_{A_j} + b_{A_j} N_E^*} \tag{19}$$

and similarly:

$$I_j^* = \frac{2\gamma r_j u_j N_j^*}{\left(r_j + d_{A_j} + b_{A_j} N_E^*\right)\left(d_{I_j} + b_{I_j} N_E^*\right)} \tag{20}$$

Considering $\dot{N}_j = 0$ we find:

$$N_j^* = \frac{1}{u_j} N_E^* \left(b_{I_j} I_j^* + b_{A_j} A_j^*\right) \tag{21}$$

$N_j^* = 0$ for all j implies both $A_j^* = 0$ and $I_j^* = 0$, and hence a steady state in which a cell-type j has zero niche-bound cells but non-zero free cells cannot exist. However, equation (21) implies existence of a trivial steady state.

Theorem 3.2: Trivial steady state S_0^*

A *trivial steady state*, denoted S_0^*, always exists. In the trivial steady state $N_j^* = A_j^* = I_j^* = 0$, $\forall j \in \{1, \ldots, n\}$.

The number of empty niches N_E^* in the trivial steady state is $N_E^* = K$ since $N_E = K - \sum_{i=1}^n N_i$.
We proceed to investigate the case $N_j^* > 0$. From equations (19), (20) and (21) we obtain:

$$N_j^* = N_E^* \left(b_{I_j} \frac{2\gamma r_j}{\left(r_j + d_{A_j} + b_{A_j} N_E^*\right)\left(d_{I_j} + b_{I_j} N_E^*\right)} + b_{A_j} \frac{1}{r_j + d_{A_j} + b_{A_j} N_E^*} \right) N_j^* \tag{22}$$

which for $N_j^* > 0$ simplifies to

$$r_j + d_{A_j} + b_{A_j} N_E^* = b_{I_j} N_E^* \frac{2\gamma r_j}{d_{I_j} + b_{I_j} N_E^*} + b_{A_j} N_E^* \tag{23}$$

since we assume $N_E^* \geq 0$. Hence:

$$N_E^* = K - \sum_{i=1}^n N_i^* = \frac{d_{I_j}\left(r_j + d_{A_j}\right)}{b_{I_j}\left((2\gamma - 1) r_j - d_{A_j}\right)}, \tag{24}$$

assuming $(2\gamma - 1) r_j \neq d_{A_j}$. Thus, requiring $N_j^* > 0$ implies a unique value of $N_E^* = K - \sum_{i=1}^n N_i^*$. For later purposes we define the *fitness* of the j^{th} clone:

Definition 3.1: Fitness

The *fitness* of the j^{th} clone is defined as:

$$F_j = \frac{b_{I_j}\left((2\gamma - 1) r_j - d_{A_j}\right)}{d_{I_j}\left(r_j + d_{A_j}\right)}. \tag{25}$$

Hence, equation (24) can be written $K - \sum_{i=1}^n N_i^* = F_j^{-1}$. Since in general $0 \leq \sum_{i=1}^n N_i \leq K$, we find that $0 \leq F_j^{-1} \leq K$ must hold. Hence, for feasibility of a steady state in which $N_j^* > 0$ we have in particular $(2\gamma - 1) r_j > d_{A_j}$ and $F_j^{-1} < K$. In the case $F_j^{-1} = K$, we would have $\sum_{i=1}^n N_i = 0$ and hence $N_j^* > 0$ is violated and only S_0^* exists.

A_j^* and I_j^* are unique and given in terms of N_j^* due to equations (19) and (20). In addition, A_j^* and I_j^* are non-zero only if N_j^* is non-zero.

Lemma 3.1: Steady state values for free cells

For any non-trivial steady state, with $N_j^* > 0$ and $K - \sum_{i=1}^n N_i^* = F_j^{-1}$, A_j^* and I_j^* are unique and given by equation (19) and (20) respectively.

Lemma 3.1 allows us to restrict our focus to just N_j^*.
A steady state where only the j^{th} clone is non-zero exists, with $N_j^* = K - F_j^{-1}$.

Theorem 3.3: Single-clone steady state S_j^*

Given $(2\gamma - 1) r_j > d_{A_j}$ and $F_j^{-1} < K$, a *single-clone steady state* exists for which $N_j^* > 0, A_j^* > 0, I_j^* > 0$ and $N_E^* > 0$ while $\forall k \neq j : N_k^* = A_k^* = I_k^* = 0$. In particular, $N_j^* = K - F_j^{-1}$ and $N_E^* = F_j^{-1}$.
We denote the unique j^{th} single-clone steady state as S_j^*.

Equation (24) holds for any clone j. This allows for existence of steady states where the multiple clones co-exist, if F_j is equal for the clones.

As a particular example, we consider a scenario with two clones, 1 and 2, with equal fitness F (i.e. $F = F_1 = F_2$). The clones have equal fitness only when

$$\frac{b_{I_1}\left((2\gamma - 1) r_1 - d_{A_1}\right)}{d_{I_1}\left(r_1 + d_{A_1}\right)} = \frac{b_{I_2}\left((2\gamma - 1) r_2 - d_{A_2}\right)}{d_{I_2}\left(r_2 + d_{A_2}\right)} \tag{26}$$

Note that the conditions $(2\gamma-1)r_1 > d_{A_1}$ and $(2\gamma-1)r_2 > d_{A_2}$ as well as $F < K$ must hold for existence of a non-trivial steady state. In a co-existences steady state, the empty niches are uniquely determined as $N_E^* = F^{-1}$. In addition, it must hold that $F^{-1} + N_1^* + N_2^* = K$ which constrains the possible values of N_1 and N_2 (and in turn A_1^*, I_1^*, A_2^* and I_2^*, due to lemma 3.1) In particular, for $N_1^* = \beta$, $N_2^* = K - F^{-1} - \beta$. Since F^{-1} is uniquely determined, this is a line parametrized by β, connecting the single-clone steady states S_1^* (when $\beta = K - F^{-1}$) and S_2^* (when $\beta = 0$).

Similar relations may be found for 3 distinct clones (Resulting in a triangle of co-existences steady states) as well as higher numbers of clones.

Theorem 3.4: Co-existence steady states

Co-existence steady states in which multiple clones assume positive concentrations may exist.
A necessary and sufficient condition for co-existence is that F_j is equal for all co-existing clones.
The number of empty niches, N_E^*, in the co-existence steady states is given uniquely by $N_E^* = F_j^{-1}$.
Additionally, the bound cells must fulfil the condition $\sum_{i\in\mathcal{C}} N_i = K - F_j^{-1}$ where $\mathcal{C}$ is the set of all co-existing clones.
This implies that co-existence steady states exists on a simplex where the dimension of the simplex is the number of co-existing clones, $\dim(\mathcal{C})$.
When there is no ambiguity, we denote the simplex of steady states S_C^*.

We note that multiple co-existence steady states can exist. As an example, consider a scenario with four clones. In such a scenario, we could have $F_1 = F_2 \neq F_3 = F_4$, which implies two separate lines of co-existence, one connecting S_1^* and S_2^* and another connecting S_3^* and S_4^*. Similarly, many different combinations of co-existence steady states can exist when two or more clones have the same fitness. However, such situations of equal fitness are unlikely to persist in a noisy biological setting. Because of this, we do not discuss the vast multitude of possible combinations of co-existence, and instead restrict our focus to situations without multiple simplexes of co-existence steady states. In some part we do however briefly comment on the case with co-existence of just two distinct clones, as it is exemplary for situations with multiple cases of equal fitness.

We sum up the feasible steady states. When the fitness of all clones are different, there are $n+1$ steady states: A trivial steady state, S_0^*, (theorem 3.2) and n single-clone steady states, S_j^* (theorem 3.3). In case of k clones with equal fitness, a k-dimensional simplex of steady states, as described in theorem 3.4.

As a particular example, the two-dimensional case features three steady states and, if $F_1 = F_2$, a line of steady states for which $N_1^* + N_2^* = K - F_1^{-1} = K - F_2^{-1}$.

Omitting A_1^*, I_1^*, A_2^* and I_2^* due to lemma 3.1, but including N_E^* for clarity, we order a subset of variables as $(N_E^*, N_1^*, N_2^*)^T$, and the steady states are written as:

$$S_0^* = \begin{pmatrix} K \\ 0 \\ 0 \end{pmatrix} \quad , \quad S_1^* = \begin{pmatrix} F_1^{-1} \\ K - F_1^{-1} \\ 0 \end{pmatrix} \quad , \quad S_2^* = \begin{pmatrix} F_2^{-1} \\ 0 \\ K - F_2^{-1} \end{pmatrix} \tag{27}$$

When $F_1 = F_2$, a line of steady states also exists:

$$S_C^* = \begin{pmatrix} F_1^{-1} \\ \beta \\ K - F_1^{-1} - \beta \end{pmatrix} \tag{28}$$

where β is a number between 0 and $K - F_1^{-1}$. We refer to this line as the co-existence steady states for cell-types 1 and 2.

In figure 2 all possible combinations of steady states for one, two and three clones are illustrated. Since the unique values of A_j^* and I_j^* follow directly from N_j^*, the figures depict only the values of N_j^*.

As we describe below, the stability of steady states can be determined. In all panels of figure 2 the white circle signifies an unstable steady state, black circles signify steady states with no positive eigenvalues while the grey circle signifies a steady state which is stable only on the single-cell subspace.

3.3 Local stability of steady states

In this section, we determine the local stability of the steady states discussed above.

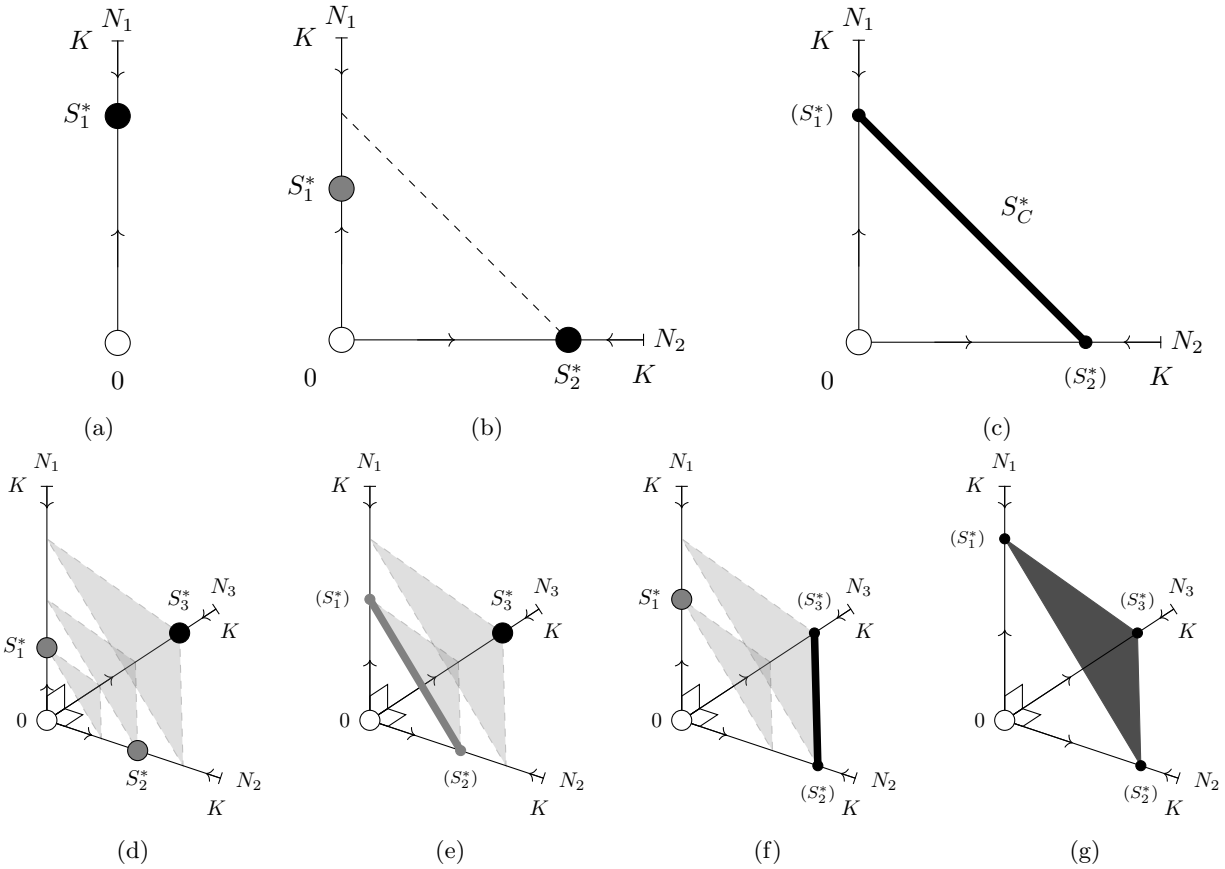


Figure 2: **Illustrations of all possible combinations of steady states for one, two and three feasible clones.** For clones that permit feasible single-clone steady states, all possible combinations are shown here. Hence this figure is a complete visual description of one-, two-, and three-clone scenarios up to permutations. As A_j^* and I_j^* follow directly from N_j^*, the figures depict the N_j-axes. In all figures, the white circle depicts the trivial steady state S_0^* in which all cell-counts are zero. Grey circles are steady states that are not locally stable, but are attracting within a single-clone subspace. Black circles and thick black lines depict locally stable steady states. Panel (a) shows a single-clone scenario. When $(2\gamma-1)r_1 > d_{A_1}$, the S_1^* steady state is feasible and stable. Panel (b) depicts a scenario with two distinct clones. In particular, $F_1 < F_2$, which implies that the S_2^* steady state is stable and both the trivial and the S_1^* steady state is unstable. If no cells from clone 2 are present, the S_1^* steady state is attracting. The dashed line is only for illustrative purposes. In panel (c), a two-clone scenario is also shown, but here $F_1 = F_2$, resulting in a line of steady states connecting S_1^* and S_2^*. Numerical investigations show that this line is attracting, however, as shown in supplementary material B, the steady states have a zero eigenvalue and hence the steady states are non-hyperbolic. Panels (d) through (g) depict the full range of possible scenarios with three clones. In (d) all three clones have different fitness. In (e) $F_1 = F_2 < F_3$, while panel (f) has $F_1 < F_2 = F_3$. Finally, panel (g) depicts the scenario where $F_1 = F_2 = F_3$. The greyed out triangles in panel (d-f) are guides for the eye, while the black triangle in panel (f) is a triangle of co-existence steady states S_C^*.

Using the formulation of equations (9), we order the $3n$ variables as $(N_1, I_1, A_1, \ldots, N_n, I_n, A_n)$. This allows us to write the Jacobian as a block-matrix of the form:

$$Jac = \begin{pmatrix} D_1 & G_1 & G_1 & \ldots & G_1 \\ G_2 & D_2 & G_2 & \ldots & G_2 \\ G_3 & G_3 & D_3 & \ldots & G_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_n & G_n & G_n & \ldots & D_n \end{pmatrix} \tag{29}$$

where

$$D_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j - u_j & b_{I_j} \left(K - \sum_{i=1}^n N_i\right) & b_{A_j} \left(K - \sum_{i=1}^n N_i\right) \\ b_{I_j} I_j & -b_{I_j} \left(K - \sum_{i=1}^n N_i\right) - d_{I_j} & 2\gamma r_j \\ u_j + b_{A_j} A_j & 0 & -b_{A_j} \left(K - \sum_{i=1}^n N_i\right) - r_j - d_{A_j} \end{pmatrix} \tag{30}$$

and

$$G_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j & 0 & 0 \\ b_{I_j} I_j & 0 & 0 \\ b_{A_j} A_j & 0 & 0 \end{pmatrix} \tag{31}$$

To determine the stability of the steady states described above, we investigate the eigenvalues of the Jacobian evaluated at the steady states. We note that the matrix D_j is the Jacobian for the subspace considering only the j'th clone, i.e. the subspace with all other N_k, I_k and A_k equal to zero, for $k \neq j$.

At the trivial steady state, S_0^*, the matrices G_j are zero, and hence the eigenvalues of the full Jacobian is given by the eigenvalues of the D_j matrices, since the Jacobian is a diagonal block-matrix:

$$Jac|_{S_0^*} = \begin{pmatrix} D_1|_{S_0^*} & 0 & 0 & \ldots & 0 \\ 0 & D_2|_{S_0^*} & 0 & \ldots & 0 \\ 0 & 0 & D_3|_{S_0^*} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & D_n|_{S_0^*} \end{pmatrix} \tag{32}$$

where $|_{S_0^*}$ denotes evaluation in S_0^*.

In the k'th single-clone steady state, S_k^*, any G_j where $j \neq k$ is zero. Thus the eigenvalues of the Jacobian are again given by the eigenvalues of the D_j matrices. For illustration, the Jacobian in the single-clone steady state with clone 1, $Jac|_{S_1^*}$, is given as:

$$Jac|_{S_1^*} = \begin{pmatrix} D_1|_{S_1^*} & G_1|_{S_1^*} & G_1|_{S_1^*} & \ldots & G_1|_{S_1^*} \\ 0 & D_2|_{S_1^*} & 0 & \ldots & 0 \\ 0 & 0 & D_3|_{S_1^*} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & D_n|_{S_1^*} \end{pmatrix} \tag{33}$$

Lemma 3.2: Stability of S_0^* and S_j^* is determined by D_j

In the trivial steady state S_0^* and in the single-clone steady states S_j^*, the eigenvalues of the Jacobian of the system of equations (9) evaluated in the given steady state values, are equal to the eigenvalues of the submatrices D_j, defined in equation (30).
Hence, the stability of the steady states can be determined by considering the submatrices D_j.

Before considering the co-existence steady states, S_C^*, we first discuss the sign of the eigenvalues of D_j.

In appendix B, we go through the process of checking the Routh-Hurwitz stability criterion for D_j. The results are summarized in the lemma below:

Lemma 3.3: Routh-Hurwitz stability criterion for D_j

Assume that S_j^* is feasible and hence $(2\gamma - 1)r_j > d_{A_j}$ as well as $K - F_j^{-1} > 0$ holds.
The matrix D_j of equation (30) evaluated in the trivial steady state, S_0^*, i.e. $D_j|_{S_0^*}$, has at least one eigenvalue with positive real part.
Evaluated in S_j^*, the j^{th} single-clone steady state, all eigenvalues of the matrix $D_j|_{S_j^*}$ have negative real part.
The eigenvalues of $D_j|_{S_k^*}$ where $k \neq j$ (i.e. the matrix with parameters of the j^{th} clone, evaluated in the single-clone steady state of the k^{th} clone) depend on the fitnesses of both the j^{th} clone and the k^{th} clone. In particular, if $F_k > F_j$, all eigenvalues have negative real part, and when $F_k < F_j$ at least one eigenvalue has positive real part.

Hence, for scenarios where no clones have equal fitness, we have determined the signs of the real parts of the eigenvalues of D_j when evaluated in S_0^* and S_j^*. This in turn determines the local stability of S_0^* and S_j^*.

Theorem 3.5: Stability of S_0^* and S_j^* for all j

If the system has a feasible single-clone steady state S_j^*, the trivial steady state S_0^* is locally unstable. Otherwise, it is locally stable.
Assuming no clones have equal fitness, a particular single-clone steady state S_k^* is stable if and only if $F_k > F_j$ holds for all $j \neq k$. If it does not hold, the steady state S_k^* is unstable.
In the special case where only one clone is considered, S_1^* is stable.

Thus, when no fitnesses are equal, the system has exactly one stable steady state and n unstable steady states. In particular, the stable steady state is the single-clone steady state of the clone with the highest fitness. Our use of *fitness* for F_j is thus justified.

Stability of the co-existence steady state

As discussed above, a simplex of solutions exists when multiple cell-types have equal fitness. As such situations are unlikely to arise in a noisy biological setting, we do not discuss them in great detail. In supplementary material B we show that in a scenario with two clones with equal fitness, the line of co-existences steady states has an eigenvalue of zero with a corresponding eigenvector along the line. Numerical investigations have shown that there is a zero eigenvalue along the simplex in general, see supplementary material B. When multiple clones have the same fitness, the Jacobian evaluated in the simplex of steady states has a zero eigenvalue. Furthermore, the numerical investigations also show that the eigenvalues depend on the fitness in the expected way: If the clones with equal fitness have a higher fitness than all other clones, the remaining eigenvalues are negative, and hence solutions of the system will approach the simplex of steady states. In the scenario where another clone has a higher fitness, the single-clone steady state of that type is stable, as expected. These relations are summarized in table 2.

In conclusion, the considerations about an equal fitness scenario shown in table 2 together with theorem 3.5 provides a complete description of hyperbolic steady states of the system and the criteria for non-hyperbolic steady states to arise.

	$F_1 > F_2 > F_3 > \ldots$	$F_1 = F_2 > F_3 > \ldots$	$F_1 = F_2 = F_3 > \ldots$	$F_1 > F_2 = F_3 > \ldots$
S_1^*	Stable	Zero eigenvalue	Zero eigenvalue	Stable
S_2^*	Unstable	Zero eigenvalue	Zero eigenvalue	Zero eigenvalue
S_3^*	Unstable	Unstable	Zero eigenvalue	Zero eigenvalue
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S_0^*	Unstable	Unstable	Unstable	Unstable
Co-existence?	No	Line (S_1^* and S_2^*)	Triangle (S_1^*, S_2^* and S_3^*)	Line (S_2^* and S_3^*)

Table 2: **Overview of the stability of steady states for examples of combinations of fitness.** All steady states are assumed to be feasible. The row labeled co-existence describes the given simplex of co-existence steady states and the points it connects. All steady states located on the co-existence simplex lead to the existence of a zero eigenvalue of the related Jacobian.

4 Model reduction

We present a model reduction. Scaling of variables together with two assumptions reduce the $3n$ dimensional system of equations (9) to an n dimensional system. The reduced system maintains most of the features and dynamics of the full system. While our considerations above yielded a complete description of the steady states of the model, a reduced model offers greater versatility as the number of parameters is reduced.

We make two general assumptions based on biological considerations:

Assumption 1 Only few niches remain empty at any given time, as found in the work of Bhattacharya et al. (2009). Hence N_E should be low in numbers compared to $N_j + I_j + A_j$ in steady state.

Assumption 2 Under most circumstances, the majority of HSC are niche-bound. Hence, in steady state, N_j are numerous compared to A_j and I_j.

Assumption 2 is in agreement with HSC mobilization studies (Yang et al., 2009) and the low division frequencies of HSC (Lee-Six et al., 2018).

Before proposing a reduced model, we first introduce a simplification by setting $b_{A_j} = 0$. Biologically, b_{A_j} is the re-attachment rate of HSC that have not yet divided after detaching from the niche. The simplification is justified, since such cells that immediately rebind can be considered included in the count of bound cells, N_j. Note that b_{A_j} does not contribute to either the steady state values N_j^* and N_E^*, or to the fitness, equation (25). The local stability of the steady states described above has been determined to be unchanged when $b_{A_j} = 0$. In addition, preliminary numerical investigations showed that the parameter should be small for the model to agree with data from literature. The model presented separately (Pedersen et al., nd) is this special case, with the additional assumption of $\gamma = 1$.

We introduce a scaling of the variables. Denoting unscaled variables with a superscript o and denoting the new scaled variables with the original notation without superscript, we define $N_j^o = \widetilde{N_j} N_j$, $I_j^o = \widetilde{I_j} I_j$ and $A_j^o = \widetilde{A_j} A_j$. Our particular choices of scalings are $\widetilde{N_j} = K$, $\widetilde{I_j} = \frac{u_j}{b_{I_j}}$ and $\widetilde{A_j} = \frac{u_j}{r_j} K$. Defining the smallest u_j as U, i.e. the u_j such that $u_j \leq u_k$ holds for all k, we introduce a scaled time-variable: $\tau = Ut$. The scaled version of equations (9) with $b_{A_j} = 0$ is thus:

$$\frac{dN_j}{d\tau} = \frac{u_j}{U}\left(\left(1 - \sum_{i=1}^{n} N_i\right) I_j - N_j\right) \tag{34a}$$

$$U\frac{dI_j}{d\tau} = 2\gamma b_{I_j} K A_j - b_{I_j} K \left(1 - \sum_{i=1}^{n} N_i\right) I_j - d_{I_j} I_j \tag{34b}$$

$$U\frac{dA_j}{d\tau} = r_j N_j - (r_j + d_{A_j}) A_j \tag{34c}$$

where N_j, I_j and A_j now denote the scaled variables. The scaling of N_j also scales the domain of functions that are solutions to $\dot{N}_j$, and hence $N_j(t) \in [0,1]$ while still $A_j(t), I_j(t) \in [0,\infty)$. The empty niches are also scaled, such that we now have $N_E = 1 - \sum_{i=1}^{n} N_i \in [0,1]$.

From the steady state considerations discussed in the previous section, we note that the steady state values of I_j^o and A_j^o scale linearly with u_j (Equations (20) and (19)), while the steady state values of N_j^o and N_E^o are independent of u_j, see theorem 3.3. In addition, the steady state value of N_E^o scales with $b_{I_j}^{-1}$. Hence, Assumption 1, stating that the steady state value of N_E^o is small compared to those of N_j^o, I_j^o and A_j^o, can be mathematically interpreted as $\frac{u_j}{b_{I_j}} \ll 1$.

We divide by $b_{I_j}K$ in equation (34b). When we defined K in the text immediately before equations (9), we assumed $K > 1$ to hold in general. Hence $\frac{U}{b_{I_j}K} \leq \frac{u_j}{b_{I_j}K} < \frac{u_j}{b_{I_j}} \ll 1$. Letting $\frac{U}{b_{I_j}K} \to 0$, we obtain a quasi-steady-state approximation for I_j for all j.

Lemma 4.1: Quasi-steady-state approximation, I_j

The expression

$$I_{j,redu} = \frac{2\gamma}{\frac{d_{I_j}}{b_{I_j}K} + 1 - \sum_{i=1}^{n} N_i} A_j \tag{35}$$

is a quasi-steady-state approximation of equation (34b), which is valid under Assumption 1.

Note that $I_{j,redu}$ scales with A_j.

Dividing equation (34c) by $r_j + d_{A_j}$, the expression $\frac{U}{r_j+d_{A_j}}$ appears on the left-hand-side. When $b_{A_j} = 0$, the steady state value of A_j^o, equation (19), is simply $\frac{u_j}{r_j+d_{A_j}} N_j^{o*}$. Under Assumption 2, N_j are numerous compared to I_j and A_j, or equivalently, $\frac{u_j}{r_j+d_{A_j}} \ll 1$. Hence $\frac{U}{r_j+d_{A_j}} \leq \frac{u_j}{r_j+d_{A_j}} \ll 1$ suggests the limit $\frac{U}{r_j+d_{A_j}} \to 0$, and a quasi-steady-state approximation for A_j arises.

Lemma 4.2: Quasi-steady-state approximation, A_j

The expression

$$A_{j,redu} = \frac{r_j}{r_j + d_{A_j}} N_j \tag{36}$$

is a quasi-steady-state approximation of equation (34c), which is valid under Assumption 2.

The two quasi-steady-state approximations together give rise to a reduced model, which is valid under Assumption 1 and Assumption 2. We first define the domain of the reduced model.

Definition 4.1: Reduced domain, $\mathbb{T}$

We define the *reduced domain* as:

$$\mathbb{T} = \{(N_1, \ldots, N_n) : N_j \geq 0, j = 1, \ldots, n \wedge \sum_{j=1}^{n} N_j \leq 1\} \tag{37}$$

By substituting A_j in equation (35) by $A_{j,redu}$ and subsequently substituting I_j in equation (34a) by the resulting expression for $I_{j,redu}$, the reduced model is obtained.

Definition 4.2: Reduced Model

On the domain $\mathbb{T}$, the *Reduced Model*, is given by:

$$\dot{N}_j = u_j \left(\frac{2\gamma\rho_j(1 - \sum_{i=1}^{n} N_i)}{\alpha_j + 1 - \sum_{i=1}^{n} N_i} - 1 \right) N_j \tag{38}$$

where $\rho_j = \frac{r_j}{r_j+d_{A_j}}$ and $\alpha_j = \frac{d_{I_j}}{b_{I_j}K}$. The variables I_j and A_j are given by equations (35) and (36) respectively.

Note that equation (38) is given in unscaled time. We refer to the parameters u_j, ρ_j and α_j as the reduced parameters. While ρ_j and α_j are dimension-free, u_j has dimension $[time]^{-1}$.

4.1 Steady states of the reduced model

The reduced form of the model, equation (38), was obtained through quasi-steady-state approximations of $\dot{I}_j$ and $\dot{A}_j$, and hence the steady states of the reduced form and the full system are the same. We here derive the steady states for illustration, and show that the stability of the steady states of the reduced form mirrors the full system. In the reduced model, I_j and A_j are obtained from N_j through equations (35) and (36) respectively, so we only consider N_j^*.

As discussed above, the niche-bound cells are scaled with K. Hence $0 \leq \sum_{i=1}^{n} N_i \leq 1$ must hold generally, and the empty niches are given as $N_E = 1 - \sum_{i=1}^{n} N_i$.

In steady state, we find that either $N_j^* = 0$ or

$$1 - \sum_{i=1}^{n} N_i^* = \frac{\alpha_j}{2\gamma\rho_j - 1} \tag{39}$$

assuming $2\gamma\rho_j \neq 1$. Note that the condition $(2\gamma - 1)r_j > d_{A_j}$ described in section 3.2, is equivalent to $2\gamma\rho_j > 1$.

As with the scaled variables, we introduce a scaling of the fitness. The scaled fitness is $f_j = KF_j$, where F_j is the fitness of the full model as defined in definition 3.1. Hence the fitness in terms of the reduced parameters ρ_j and α_j is:

$$f_j = \frac{2\gamma\rho_j - 1}{\alpha_j} \tag{40}$$

and hence equation (39) can be written as $1 - \sum_{i=1}^{n} N_i^* = f_j^{-1}$.

This leads to the reduced forms of the steady states.

Theorem 4.1: Trivial steady state S_0^*, reduced form

A *trivial steady state*, denoted S_0^*, always exists. In the trivial steady state $N_j^* = 0$, $\forall j \in \{1, \ldots, n\}$.

When $0 < f_j^{-1} < 1$, a single-clone steady state exists and is feasible.

Theorem 4.2: Single-clone steady state S_j^*, reduced form

Given $2\gamma\rho_j > 1$ and $f_j^{-1} < 1$, a *single-clone steady state* exists for which $N_j^* > 0$ while $\forall k \neq j : N_k^* = 0$
In particular, $N_j^* = 1 - f_j^{-1}$ and $N_E^* = f_j^{-1}$.
We denote the j^{th} single-clone steady state as S_j^*.

As for the full model, for clones with equal fitness, co-existence steady states also exist, with the condition $\sum_{i \in \mathcal{C}} N_i = 1 - f^{-1}$ where $\mathcal{C}$ is the set of all co-existing clones and f is the fitness for the clones.

Two-clone example

For illustration, the steady states of the two-clone reduced form, ordered as $(N_1^*, N_2^*)^T$, are:

$$S_0^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad , \quad S_1^* = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix} \quad , \quad S_2^* = \begin{pmatrix} 0 \\ 1 - f_2^{-1} \end{pmatrix} \tag{41}$$

If $f_1 = f_2$, a line of steady states exists, given as:

$$S_C^* = \begin{pmatrix} \beta \\ 1 - f^{-1} - \beta \end{pmatrix} \tag{42}$$

where $f = f_1 = f_2$ and $\beta \in [0, 1 - f^{-1}]$. Note that both S_1^* and S_2^* are on the line, with $\beta = 1 - f^{-1}$ and $\beta = 0$ respectively.

Local stability of the steady states of the reduced model

To investigate the stability of the steady states of the reduced form, we first define:

$$g_j(N_1, \ldots, N_n) = u_j \alpha_j f_j \left(\frac{1 - \sum_{i=1}^{n} N_i - f_j^{-1}}{\alpha_j + 1 - \sum_{i=1}^{n} N_i} \right) \tag{43}$$

and

$$h_j(N_1, \ldots, N_n) = -2\gamma\rho_j\alpha_j u_j \left(\frac{1}{\alpha_j + 1 - \sum_{i=1}^{n} N_i} \right)^2 N_j \tag{44}$$

As the functional dependence on $N_1, \ldots, N_n$ is clear, we simply write g_j and h_j for the rest of this section.

Note that equation (38) can be written as $\dot{N}_j = g_j N_j$ and that $\frac{\partial g_j}{\partial N_j} = h_j N_j^{-1}$

This allow us to write the Jacobian for the reduced system as:

$$Jac_{redu} = \begin{pmatrix} h_1 + g_1 & h_1 & \ldots & h_1 \\ h_2 & h_2 + g_2 & \ldots & h_2 \\ \vdots & \vdots & \ddots & \vdots \\ h_n & h_n & \ldots & h_n + g_n \end{pmatrix} \tag{45}$$

In the trivial steady state, S_0^*, we find

$$g_j|_{S_0^*} = \frac{\alpha_j u_j}{\alpha_j + 1} (f_j - 1) \tag{46}$$

for all j. The feasibility criteria for the single-clone steady state S_j^*, implies that $f_j > 1$ when S_j^* is feasible. Thus the sign of g_j in the trivial steady state depends on the feasibility of the single-cell steady state.

Evaluating in the j'th single-clone steady state S_j^* we find $g_j|_{S_j^*} = 0$. For any $k \neq j$, we find that:

$$g_k|_{S_j^*} = \frac{u_k}{\alpha_k + \frac{1}{f_j}} \left(\frac{f_k}{f_j} - 1 \right) \tag{47}$$

Hence the sign of g_k is determined by the relative fitness of the two clones. When clone k is more fit than clone j, $g_k|_{S_j^*} > 0$, and when they have the same fitness, it is zero.

We note that $N_j = 0 \Rightarrow h_j = 0$. Thus $h_j|_{S_0^*} = 0$, as well as $h_k|_{S_j^*} = 0$ for any $k \neq j$. In addition, when $N_j^* > 0$ we find that $h_j < 0$, and hence $h_j|_{S_j^*} < 0$ as well as $h_j|_{S_C^*} < 0$.

Evaluating the Jacobian in the trivial steady state we find:

$$Jac_{redu}|_{S_0^*} = \begin{pmatrix} g_1|_{S_0^*} & 0 & \ldots & 0 \\ 0 & g_2|_{S_0^*} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g_n|_{S_0^*} \end{pmatrix} \tag{48}$$

Since the matrix is diagonal, the eigenvalues are simply $g_j|_{S_0^*}$, and hence the stability of S_0^* is unstable if any $f_j > 1$, and stable only if for all j, $f_j < 1$, in which case S_0^* is the only feasible equilibrium.

The stability of the single-clone steady state S_j^* follows that of the full model. For illustration, we evaluate the Jacobian in the steady state for the 1^{st} cell-type:

$$Jac_{redu}|_{S_1^*} = \begin{pmatrix} h_1|_{S_1^*} & h_1|_{S_1^*} & \ldots & h_1|_{S_1^*} \\ 0 & g_2|_{S_1^*} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g_n|_{S_1^*} \end{pmatrix} \tag{49}$$

The matrix is upper triangular and hence the eigenvalues are again easily found. For any $j \neq 1$, the sign of $g_j|_{S_1^*}$ is determined by $\frac{f_j}{f_1} - 1$. When $g_j|_{S_1^*} < 0$, $f_1 > f_j$. Since $h_1|_{S_1^*} < 0$ always, we note that S_1^* is stable when $f_1 > f_j$ holds for all j, and unstable if any $f_j > f_1$. This result holds for any single-clone steady state.

Finally, in a co-existence steady state, S_C^*, where m clones have the same fitness and p clones have lower fitness, the Jacobian has m rows of $h_j|_{S_C^*} < 0$ and p rows of zeros. As an example, a three dimensional case where $f_1 = f_2 > f_3$ we have:

$$Jac_{redu}|_{S_C^*} = \begin{pmatrix} h|_{S_C^*} & h|_{S_C^*} & h|_{S_C^*} \\ h|_{S_C^*} & h|_{S_C^*} & h|_{S_C^*} \\ 0 & 0 & 0 \end{pmatrix} \tag{50}$$

where $h|_{S_C^*} = h_1|_{S_C^*} = h_2|_{S_C^*}$, and hence the eigenvalues are $2h|_{S_C^*} < 0$ and 0 (with algebraic multiplicity 2).

In conclusion, the stability of the reduced model mirrors that of the full model: Fitness of clones determines the stability in such a way that the single-clone steady state for the clone with the highest fitness is stable.

4.2 Change of variables and further simplifications

Biologically, the relative frequency of a malignant clone can be considered a measure for the molecular residual disease in the stem cell compartment.

On the set $\mathbb{T} \setminus S_0^*$ in $\mathbb{R}^n$, we define the sum of total niche-bound cells:

$$T = \sum_{i=1}^{n} N_i \tag{51}$$

and the relative frequency of a clone for $j \in (2, 3, \ldots, n)$ as:

$$C_j = \frac{N_j}{\sum_{i=1}^{n} N_i} \tag{52}$$

Note that we can calculate C_1 as $C_1 = 1 - \sum_{i=2}^{n} C_i$. As the trivial steady state $S_0^* = \{(0, \ldots, 0)\}$ has been omitted, $T > 0$ holds and C_j is well-defined.

The n differential equations for the reduced model in equation (38) can be given in terms of T and C_j.

Definition 4.3: Transformed Model

On the domain $(0, 1] \times [0, 1]^{n-1}$ the *transformed model* is given as:

$$\dot{T} = T \sum_{i=1}^{n} g_i(T) C_i \tag{53a}$$

$$\dot{C}_j = \left(g_j(T) \sum_{i=1}^{n} C_i - \sum_{i=1}^{n} g_i(T) C_i \right) C_j \text{ for } j = 2, \ldots n \tag{53b}$$

$$C_1 = 1 - \sum_{i=2}^{n} C_i \tag{53c}$$

where $g_j(T)$ is given as

$$g_j(T) = u_j \alpha_j f_j \left(\frac{1 - T - f_j^{-1}}{\alpha_j + 1 - T} \right) \tag{54}$$

Lemma 4.3: Equivalence of the reduced model and the transformed model

When excluding the trivial steady state S_0^* from the domain of the reduced model, the transformed model is equivalent to the reduced model from definition 4.2 in the sense that solutions of one model can be transformed to equivalent solutions of the other model.

The function

$$\mathcal{F} : (N_1, \ldots, N_n) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n \mapsto (T, C_2, \ldots, C_n) = \left(\sum_{i=1}^{n} N_i, \frac{N_2}{\sum_{i=1}^{n} N_i}, \ldots, \frac{N_n}{\sum_{i=1}^{n} N_i} \right) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n \tag{55}$$

transforms points in the domain of the reduced model to the domain of the transformed model. The inverse function is trivially found as:

$$\mathcal{F}^{-1} : (T, C_2, \ldots, C_n) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n \mapsto (N_1, \ldots, N_n) = \left(T(1 - \sum_{i=2}^{n} C_i), TC_2, \ldots, TC_n \right) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n \tag{56}$$

The transformation $\mathcal{F}$ and its inverse are bijective, and hence lemma 4.3 is proven.

The function $g_j(T)$ in equation (54) determines the sign of $\dot{T}$ and $\dot{C}_j$. $g_j(T) = 0$ only when $T = T_1^* = 1 - f_j^{-1}$, i.e. the steady-state value of N_j in the single-clone steady state, S_1^*. For $T > 1 - f_j^{-1}$, then $g_j(T) < 0$ and

conversely $g_j(T) > 0$ when $T < 1 - f_j^{-1}$. Hence, for $C_j = 1$ and $C_i = 0$ for all $i \neq j$, we find that T will approach the j'th single-clone steady state:

$$T \underset{t\to\infty}{\longrightarrow} 1 - f_j^{-1} \tag{57}$$

In the transformed model, the steady states previously discussed correspond to:

$$\overline{S}_1 = \begin{pmatrix} 1-f_1^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \overline{S}_2 = \begin{pmatrix} 1-f_2^* \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \overline{S}_3 = \begin{pmatrix} 1-f_3^* \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ \overline{S}_n = \begin{pmatrix} 1-f_n^* \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{58}$$

where $\overline{S}_j$ describes $(T, C_2, C_3, \ldots, C_n)$.

Co-existence steady states also exists under the same conditions as before the transformation. We restrict our attention to a two-clone scenario where the dynamics and the co-existence steady states are clearly illustrated.

4.2.1 Total cell-count and relative HSC frequency in a two-clone scenario

While multiple populations of distinct malignant clones can arise during some hematologic malignancies, we now consider malignancies with just one population of malignant stem cells.

Such malignancies can be described as a two-clone scenario, where the one clone models the healthy HSC population and the second clone models the population of malignant stem cells. In a two-clone subsystem, we write C_2 as C to be concise (and hence $C_1 = 1 - C_2$). Equations (53a) and (53b) can be written as:

$$\dot{T} = [g_1(T)(1-C) + g_2(T)C]\,T \tag{59a}$$

$$\dot{C} = (g_2(T) - g_1(T))\,(1-C)C \tag{59b}$$

We refer to this model as the 2-*clone transformed model*. In section C of the supplementary material, a detailed analysis of the model is presented. We here summarize the most important results and refer the reader to the supplementary material for details.

For equal fitness, $f = f_1 = f_2$, the functions g_1 and g_2 are identical only if all parameters are equal (i.e. $u_1 = u_2$, $\alpha_1 = \alpha_2$ and $f_1 = f_2$). In this particular case, the system reduces to $\dot{C} = 0$ and $\dot{T} = g(T)T$ where $g = g_1 = g_2$.

For $f = f_1 = f_2$ but $\alpha_2 \neq \alpha_1$, the system has a line of steady state for which $T = 1 - f^{-1}$ and $C \in [0, 1]$.

Theorem 4.3: Transformed two-clone system, equal fitness

For $f = f_1 = f_2$, a line of steady states $\overline{S}_C = \{(T, C) = (1 - f^{-1}, \eta)\}$ for all $\eta \in [0, 1]$ exists. The line of steady states is attracting along the T-axis for all C. $\dot{C} = 0$ for all T only if $\alpha_1 = \alpha_2$ and $u_1 = u_2$. Otherwise $T_e^* = 1 + \alpha_1\alpha_2\frac{u_2 - u_1}{u_2\alpha_2 - u_1\alpha_1}$ exists and is a nullcline for $\dot{C}$, if $0 < T_e^* < 1$.

Considering now only $f_1 \neq f_2$. The steady states $\overline{S}_1$ and $\overline{S}_2$ exist, corresponding to S_1^* and S_2^* respectively. In order (T, C) they are:

$$\overline{S}_1 = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix}, \ \overline{S}_2 = \begin{pmatrix} 1 - f_2^{-1} \\ 1 \end{pmatrix} \tag{60}$$

These two steady states are the only steady states of the system when $f_1 \neq f_2$, and the global dynamics can be determined.

Theorem 4.4: Global dynamics of the transformed two-clone system

For $f_1 \neq f_2$, the system of equations (59) has just two steady states, $\overline{S}_1$ and $\overline{S}_2$, as defined in equations (60). There are no periodic solutions.
For $f_1 > f_2$, all solutions with $C(0) < 1$ are attracted toward $\overline{S}_1$.
For $f_1 < f_2$, all solutions with $C(0) > 0$ are attracted toward $\overline{S}_2$.

Details and proof of theorem 4.4 are given in section C of the supplementary material.
Nullclines for $\dot{T}$ and $\dot{C}$ can be determined.

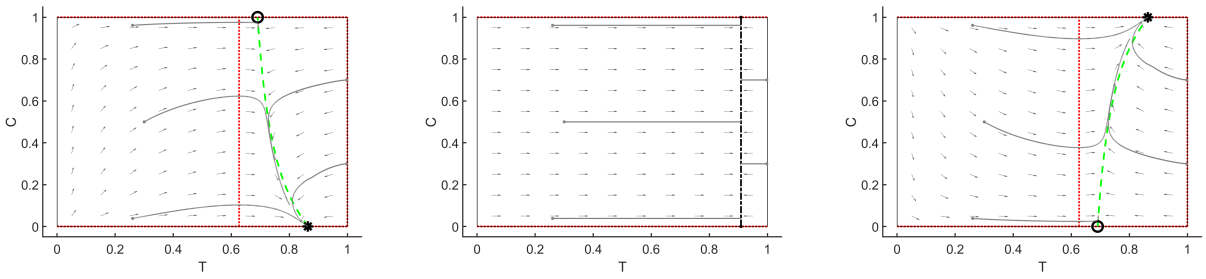
Lemma 4.4: Nullclines for the $2D$ transformed model

$\dot{C}$ has 2 to 4 nullclines within the domain. These are $C = 0$, $C = 1$, and the two values $T = T_{\pm}$ given in supplementary material C.
$T_{\pm}$ depend on parameters and are not considered nullclines when they are outside the domain.
$\dot{T}$ has a nullcline described by a value T between $1 - f_1^{-1}$ and $1 - f_2^{-1}$, and a corresponding value of $C_{null}(T)$ given as:

$$C_{null}(T) = \frac{g_1(T)}{g_1(T) - g_2(T)} \tag{61}$$

The $\dot{C}$-nullclines and the $\dot{T}$-nullcline only cross in $\overline{S}_1$ and $\overline{S}_2$.

In figure 3 some exemplary plots of the (T, C) phase-plane are shown. The nullcline(s) for $\dot{C}$ are shown in red, while the nullcline for $\dot{T}$ is shown in green. Further examples are shown in supplementary C. The figures highlight changes in the relative frequency of the HSC belonging to the two clones which could be difficult to notice when considering solutions of the reduced model for two clones. In particular, one of the simulated scenarios shown in figure 3c starts from initial conditions where $T(0) = 0.3$ and $C(0) = 0.5$. In this scenario, cell counts of both clones, N_1 and N_2, increase initially, with N_1 increasing faster. By modelling the relative frequency directly, we see that a decrease in relative frequency is followed by an increase when T approaches ≈ 0.65. Hence even if N_1 is increasing faster than N_2, the 1^{st} clone is eventually outcompeted by the 2^{nd} clone. If the 2^{nd} clone is a malignant clone, such scenarios could have clinical significance as they could arise following treatment.



(a) $f_1 > f_2$ (b) $f_1 = f_2$ (c) $f_1 < f_2$

Figure 3: **Phase-planes displaying exemplary scenarios of the 2-clone transformed model.** The open circle denotes unstable steady states while the asterisk denotes the stable steady states. Grey lines are numerical solutions, with the same initial conditions in the three panels. The green dashed lines are the nullclines for T, while the red dotted lines are the nullclines for C. The arrows show the directional field of flows. As the flow and the values of the steady states depend on the specific parameter values, the figures are examples of typical behaviours, but differences in parameter will influence the specific flows. In panel (b) the fitness is equal, and the line of co-existence steady states, $\overline{S}_C$ is shown as a black dash-dotted line

4.2.2 Additional simplification of a two-clone scenario

We introduce a final simplification of the model, valid close to $\overline{S}_1$. Evaluating in $\overline{S}_1$, i.e. $T = T_1^*$, we see from equation (54), $g_1(T_1^*) = 0$ and

$$g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right) \tag{62}$$

Assuming $T_1^* \approx T_2^*$, we approximate $\dot{T} = 0$. For $T(0) = T_1^*$, a simplified expression for the relative frequency of the HSC belonging to the two clones arises, in which the expression for $g_2(T_1^*)$ above determines the model behaviour.

Definition 4.4: Logistic Approximation describing the relative frequency of the mutated clone

The *Logistic Approximation Model* is given as

$$\dot{C} = \phi(1 - C)C \tag{63}$$

where $\phi = g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right)$. The sum of HSC is constant with $T = 1 - f_1^{-1}$.

Note that ϕ is not a dimension-free parameter and has units of $[time]^{-1}$, like u_j.

The logistic approximation model is valid close to $\overline{S}_1$, and when $T_1^* \approx T_2^*$, i.e. when the steady-state cell-count of second clone is similar to the cell-count of the first clone in steady state. This is the case when $f_1 \approx f_2$. Hence, we assume that the initial stages of a malignancy which only has a minor fitness advantage are described accurately by the logistic approximation model.

Equation (63) is a logistic equation, with steady states $C = 0$ and $C = 1$, corresponding to $\overline{S}_1$ and $\overline{S}_2$ respectively. ϕ determines the direction of flow and hence the disease progression. When $\phi > 0$, the steady state $C = 1$ is stable, while $C = 0$ is stable for $\phi < 0$. The sign of ϕ depends solely on $\frac{f_2}{f_1}$. When $f_2 > f_1$, $\phi > 0$, while $f_2 < f_1$ implies $\phi < 0$. Thus the asymptotic behaviour of the logistic approximation is in agreement with previous analysis of the model.

ϕ does not depend on u_1. This is an artefact of the approximation. When u_1 and u_2 are much different, only little agreement is found with numerical solutions of equations (59).

The logistic approximation model represents the simplest formulation that maintains the asymptotic behaviour of the model. Further simplification, e.g. to exponential growth during initial disease stages, fails to capture both $\overline{S}_1$ and $\overline{S}_2$ and the dependence on fitness as determinant of steady-state stability.

5 Model application to clinical data

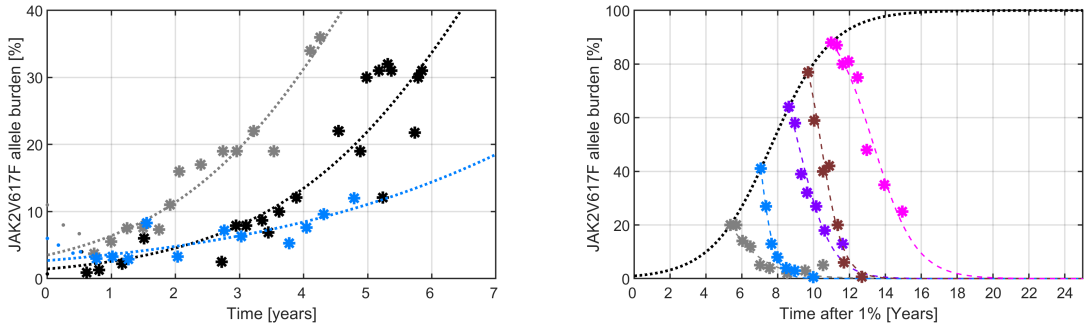
Consider a specific hematopoietic malignancy; the group of Philadelphia-negative myeloproliferative neoplasms (MPNs), characterized by the JAK2^{V617F} mutation. The allele burden of cells with the JAK2^{V617F} mutation provides an estimate of the residual disease and can be interpreted as the relative frequency of the malignant clone in our model.

We introduce examples of patient-data, previously presented (Pedersen et al., 2020). The patient-data consists of measurements of the JAK2^{V617F} allele burden identified in blood-cells during treatment-free periods and periods with mono-therapy with pegylated Interferon-α. We consider the measured JAK2^{V617F} allele burden indicative of the frequency of mutated HSC among all HSC.

A general rule for applying models to clinical practice is to keep the models as simple as possible without losing the agreement with data to avoid overparametrization. From visual inspection, the data for both the treatment-free disease progression and the development during therapy is approximately logistic. We fit the logistic approximation from equation (63) to data, by minimizing the mean square error, using the `MATLAB R2018b` function `fminsearch`. In figure 4a depicts data from three untreated MPN-patients along with fits of the logistic equation. In addition, we also fit the initial value of C. Note that the first 200 days are excluded from the fit. This is done since all three patients previously received treatment and a continued decline immediately after treatment-stop is expected (Pedersen et al., 2020).

Figure 4b displays five patients with a range of different initial JAK2^{V617F} allele burdens, receiving treatment with pegylated interferon alfa-2a (*Pegasys*) or pegylated interferon alfa-2b (*PegIntron*). The figure also shows a fit of the logistic approximation model to the data. Baseline-measurement have been time-shifted to coincide with a growth of the JAK2^{V617F} allele burden of $\phi = 0.6$, similar to two of the patients shown in figure 4a.

The data and fits presented exemplify the range of values of ϕ for disease progression and treatment for MPN-diagnosed patients.



(a) (b)

Figure 4: **Fit of the logistic approximation to data from MPN-diagnosed patients.** Panel (a) depicts data from three patients during a treatment-free period. Note that data during the initial 200 days were excluded from the fit, as delayed responses from previous treatment are expected to influence this period. For the fits, both the initial value of C and the value of ϕ was fitted to the data. Panel (b) shows a typical disease progression in the model in dotted black, along with data from five patients receiving treatment with interferon-α, time-shifted to coincide with the growth. At the time of treatment initialization, the value of ϕ is changed. The new value of ϕ was fitted such that it minimized the least-square error between the model and the given patient. In units of $years^{-1}$, the three patient-specific values of ϕ in panel (a) during untreated development of MPNs are 0.59, 0.63 and 0.30. For the development during treatment, the five patient-specific values of ϕ in panel (b) are -0.66, -0.85, -1.1, -1.7 and -2.3.

6 Discussion and conclusion

Despite the vital role of HSC in the human body, many questions about their dynamics remain unanswered. The elusive processes within the bone-marrow microenvironment have yet to be fully identified and determined quantitatively. We propose a mechanism-based mathematical model of HSC behaviour based on the state-of-the-art biological knowledge about HSC behaviour. Through mathematical analysis and numerical simulations, the model can help to clarify the general understanding of HSC dynamics, and in future work be used to generate testable hypotheses that can elucidate the open questions surrounding HSC, both in health and in disease. Through a series of considerations and assumptions based on biological knowledge, multiple model simplifications arise. This leads to a *reduced* model (Definition 4.2), a *transformed* model (Definition 4.3) and a *logistic approximation* model (Definition 4.4), in addition to the *full* model.

Mathematical analysis of a proposed model of HSC dynamics

The full form of the proposed model consists of a system of ordinary differential equations, describing HSC-specific niches in the bone-marrow and n distinct clones of HSC. In the model, HSC can be in three distinguishable states; niche-bound, active and inactive. The n clones model separates subpopulations of HSC where cells pertaining to a given clone are assumed to be identical. Considering multiple clones in the model hence corresponds to scenarios where cellular heterogeneity is present in the HSC population, such as in the case of hematologic malignancies as e.g. acute myeloid leukemia or MPN. HSC are modelled to bind reversibly to the niches. Following detachment from the niches, the HSC are in the active state. The self-renewal potential of this active state decreases during divisions outside the niche. After a certain number of divisions, cells enter an inactive state and have to rebind to the niche or differentiate. During rebinding to the niche, the HSC restore their self-renewal potential and can become reactivated. The switching between the niche-bound HSC state, the active state (outside the niche, prone to divide) and inactive state (outside the niche, prone to differentiate or reattach) is the novelty of the proposed model.

A special case of the model has previously been investigated with focus on certain clinical consequences of the modelled mechanisms (Pedersen et al., nd). In contrast, the present work is a thorough mathematical analysis of a more general model.

The proposed model bears similarities to models of HSC found in the literature. Ashcroft et al. (2017) proposed a similar model of HSC dynamics and interaction with the niches, in which detachment from the niche

was also explicitly considered. The detachment was related to a subsequent release into the peripheral blood, related to an enhanced probability of stem cell death, and not related to stem cell division in the way we consider in our model. Instead, Ashcroft et al. (2017) models HSC division such that one cell maintains its position within the niche and the other enters either an empty niche or the peripheral blood. By modelling detachment from the niche as a prerequisite for HSC division, we were able to investigate which effect such detachment could have on clonal competition. Interestingly, our model implies that clonal dominance was independent of detachment rate, which suggests that detachment in the way we modelled it only plays a minor role in competition between clones. We emphasize that our suggestion of detachment as a prerequisite for division is a novel hypothesis, and requires confirmation by experimental evidence. In the work of Stiehl et al. (2020), a model of the HSC-niche system is also considered. Similarly, cell-division occurs in the niche and one of the resulting cells maintains its position in the niche. The model does not consider explicit detachment from the niche, instead free stem cells can dislodge bound stem cells from the niche leading to a cell-cell competition for specific niches. HSC differentiate after a certain number of unsuccessful attempts to occupy a niche. Hence, the model proposed by Stiehl et al. (2020) provides a biological explanation for why HSC differentiate which is lacking in our proposed model. However, a model extension with multiple stages of free cells, similar to the extension discussed in supplementary material A, could allow for a similar interpretation in our model, by modelling an increasing probability for differentiation following each division.

Becker et al. (2019) considers a model of HSC interaction with the niche. Similar to our model, niche-bound cells are considered quiescent and free cells are proliferating. In the model described by Becker et al. (2019), HSC feedback on niche cells plays a significant role, modelling a dynamic relation between HSC and niche cells. In contrast, we implicitly assume that the number of niches are constant in our model, and instead focus on the dynamics of HSC under this condition. Modelling changes in the number of niches could be considered by a more elaborate expression for the dynamics of empty niches, the processes considered by Becker et al. (2019) could be included on our model. However, due to the uncertainty about the exact cells that make up the niche, considering such specifics as reproduction and death rates of the abstract niches we consider would be difficult. Combining the findings of our model with those of Becker et al. (2019), suggest that the biological HSC-niche system is robust, and can return to homeostasis after perturbations of both cell- and niche-count.

Analysis of our model identified an attractive trapping region. Solutions to the system of differential equations with initial conditions outside this region were found to enter the region in finite time. This implies that the model is biologically reasonable and appropriate even when the initial state is perturbed.

Mathematical analysis revealed the equilibria of the system, as well as their local stability. The existence, non-negativity and stability of the steady states were found to depend on an expression consisting of clone-specific parameters, which we denote as the HSC fitness, F_j for the j^{th} clone. A trivial steady state, S_0^*, with no HSC and only empty niche-space always exists. For each clone considered, a single-clone steady state, S_j^*, exists in which the given clone has non-zero cell-counts and all other clones are vanishing. The j^{th} clone has positive cell-counts only when the respective fitness F_j is positive. When any F_j is positive, the trivial steady state is unstable. The single-clone steady state of the clone with the higher fitness is locally stable. Determining the fitness of clones thus reveals which clone is dominant in a competitive setting. If multiple clones have the same fitness, a simplex of steady states exists in which the clones co-exists. Although such situations of co-existence could be mathematically possible, it is biologically unlikely that equal fitness is attainable for extended lengths of time due to the noisy biological environment and other systemic factors. For this reason, the co-existence of steady states can be considered a mathematical artefact rather than a biological relevant feature of the system. The complete local stability considerations of the full $3n + 1$ dimensional model are described in section 3.3.

The stability of the steady states reveals that the fitness can be used to predict the outcome of competition between HSC clones. A given clone outcompetes a less fit clone, leading to extinction of the less fit clone. Comparing this dynamic to dynamics known from ecology sheds light on the nature of competing stem cells as not simply cells that interact freely with the bone-marrow niches, but instead as populations competing for a shared resource, similar to ecological systems. The fitness expression depends solely on the properties of a given clone and its interaction with the bone-marrow niches and thus it is independent of the properties of other clones. For this reason, the fitness of a clone can be determined from single-clone experiments and the result of competition can be predicted without experiments in which multiple clones are present simultaneously. The fitness is independent of the number of niches, K. This implies that the presence of many or few niches has no qualitative effect on the asymptotic behaviour; the clone with the higher fitness will out-compete other clones regardless of niche-counts. It is possible that this independence of the number of niches is a consequence of the assumption that K is constant in the proposed model. Changes to the niche-count K could arise due to e.g. age-related growth of a young individual, a specific HSC-niche interaction as suggested by Becker et al. (2019), or feedback from the immune system in response to hematologic malignancy. Such changes could affect the entire bone-marrow microenvironment, either indirectly by changing niche-counts, or indirectly through other means. This could in turn affect the proposed model and the fitness expression. By considering HSC in isolation from external factors, the proposed model provides an estimate of HSC dynamics based on the process

parameters that we considered most significant.

Biological assumptions imply a model reduction

Based on the biological assumptions that only few niches are empty in homeostasis and that the majority of HSC are bound to the niche, a reduced model is proposed. These assumptions are interpreted mathematically and imply that quasi-steady-state approximations of non-niche-bound HSC are appropriate. For n clones, the full model has $3n$ dynamic variables, whereas the reduced model is given by an n-dimensional system of ODEs, and reduce the number of parameters considered from $6n + 2$ to $3n + 1$. The steady states of the reduced model correspond to the steady state of the full model, and stability considerations reveal that the criteria for stability of the steady states depend on HSC fitness in the same way as in the full model. In the reduced model, parameters are grouped together in aggregated parameters. Thus, it is possible that specific changes to the parameters are not captured by the reduced model, e.g. if the attachment and differentiation of inactive HSC increase equally, the reduced model remains unchanged while minor changes appear in the full model. For comparing the model with experimental or clinical data however, this implies that there can be types of data where the reduced model is identifiable but the full model is not.

Transformation of the model highlights clinical relevance

We presented a transformation of the reduced model to the total HSC cell-count and a measure of the fraction of cells pertaining to a given clone. Under the transformation the trivial steady state is omitted, however the existence and stability of the remaining steady states are identical to those in the full model. The transformed variables can be related to clinical measurements. In particular, many hematologic malignancies arise from a single malignant clone in a setting where the remaining HSC are similar enough in properties to be considered identical. As such, considering the transformation for a two-clone scenario has significant clinical relevance.

The transformed model offers a different way to understand and illustrate HSC dynamics compared to the reduced model. One particular concept with clinical relevance was illustrated in figure 3c. In a scenario where a malignant clone had a higher fitness compared to healthy HSC, we observed that for low initial counts of HSC, it was possible for the relative frequency of the malignant clone to decrease initially before approaching the attracting steady state corresponding to full disease. Clinically, this is interesting as low HSC counts can arise due to treatment such as chemotherapy. Hence, the model suggests that after chemotherapy, a patient could experience a temporary decrease in disease burden, followed by an increasing disease burden. Immediately following the treatment, it could hence wrongly be interpreted that the treatment leads to long-lasting decrease of disease burden, and give false hopes for eradication of the malignancy. Even for scenarios with equal fitness, chemotherapy leading to an equal decrease for all clones can cause changes in relative frequency, as illustrated in supplementary figure A2. It is unclear how these observations apply to clinical practice, but our findings suggest evaluating the response to treatment can be nonlinear and complex.

Further simplification allows for simple comparison to clinical data

A further simplification of the transformed model for two clones may be illustrative as well as applicable in the clinical setting. Under the assumption that HSC counts in steady state do not vary significantly between the healthy state and the full-blown disease state, the transformed model simplifies to a logistic equation for the relative frequency of the malignant clone. A single aggregated parameter, ϕ, determines the dynamics of the relative frequency, revealing how independent parameters influence the general behaviour. The steady states of the logistic approximation were the same as in the full model. Local stability of the steady states was similarly determined sby HSC fitness, and the criteria for clonal dominance in the logistic approximation model agree with the full model. The simplification offers an explanation for the success and frequent use of logistic or exponential models to describe hematologic malignancies. Our approach provides a theoretical foundation for otherwise empirical models of disease burden growth and decay.

We identified the parameter ϕ of the logistic equation for the relative frequency for an exemplary subset of a published MPN-patient cohort. Both treatment-free disease progression and periods of interferon-α monotherapy were found to agree well with the logistic approximation. Hence, the model provides an estimate for the development of the disease burden for the patients investigated. The clinical data consisted of measurements of the JAK2^{V617F} allele burden in blood cells. Although the disease burden of blood cells and HSC can be different in principle, evidence suggests that specifically for the JAK2^{V617F} mutation, the disease burdens in mature and immature cell compartments are strongly correlated (Takahashi et al., 2013). For this reason, we assumed the relative frequency of malignant cells to be equal for blood cells and HSC.

The single-parameter logistic expression agreed well with data of MPN patients during and following interferon-α treatment and set a lower bound for how well a more complex model could fit to data. When considering the scenario with two distinct HSC clones, the reduced model has 7 parameters and the full model has a total of 14 parameters. Hence, a fit of either model to the sparse data presented could result in significant

overparametrization. Thus, even if the logistic approximation appeared simplistic, it provided a reasonable estimate for HSC dynamics. Further investigation of how the different forms of the model compares to data may suggest novel *in vitro* or mice experiments to understand and directly quantify the HSC-processes that the model was based on. This could include various HSC labelling assays which could identify properties such as e.g. how many times an active HSC divides before exhaustion (γ) or if labelled cells always divide before attaching to the niche *in vivo*, implying $b_{A_j} = 0$.

Concluding remarks

In conclusion, mathematical analysis of a proposed model of hematopoietic stem cells allowed us to investigate the dynamics of a complex biological system which is inherently difficult to observe experimentally. Through a sequence of model reductions and simplifications based on biological knowledge, a simple logistic expression arose. Long-term persistence of a clone was determined by an expression of stem cell fitness, and criteria for clonal dominance were the same for the reduced and simplified models. By comparing the simplified model to clinical data, the model was found to be in agreement with clinical observations. Our results have clinical implications about hematopoietic stem cells when the various systems of feedback from the blood or immune system are not considered. In this work, we assumed that the effect of such feedbacks had little significance. How these feedbacks, such as e.g. inflammation of the bone-marrow, change the dynamics of the hematopoietic stem cells is an important topic to be investigated in future work.

7 Acknowledgements

This work was initiated during a visit of RKP at Heidelberg University. The authors would like to thank Anna Marciniak-Czochra (Heidelberg University, Germany) for supporting the realization of the research stay. RKP is grateful for the support of the "Produktion, Forskning og Innovation (PFI)" grant from Region Sjælland, Denmark.

Declarations of interest: none

References

Altrock, P. M., Liu, L. L., and Michor, F. (2015). The mathematics of cancer: Integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745.

Andersen, M., Sajid, Z., Pedersen, R. K., Gudmand-Hoeyer, J., Ellervik, C., Skov, V., Kjær, L., Pallisgaard, N., Kruse, T. A., Thomassen, M., Troelsen, J., Hasselbalch, H. C., and Ottesen, J. T. (2017). Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. *PLOS ONE*, 12(8).

Ashcroft, P., Manz, M. G., and Bonhoeffer, S. (2017). Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments. *PLOS Computational Biology*, 13(10).

Becker, N. B., Günther, M., Li, C., Jolly, A., and Höfer, T. (2019). Stem cell homeostasis by integral feedback through the niche. *Journal of Theoretical Biology*, 481:100–109.

Bhattacharya, D., Czechowicz, A., Ooi, A. L., Rossi, D. J., Bryder, D., and Weissman, I. L. (2009). Niche recycling through division-independent egress of hematopoietic stem cells. *The Journal of Experimental Medicine*, 206(12):2837–2850.

Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P., and Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells in vivo. *Blood*, 117(17):4460–4466.

Clapp, G. and Levy, D. (2015). A review of mathematical models for leukemia and lymphoma. *Drug Discovery Today: Disease Models*, 16(1):1–6.

Colijn, C. and Mackey, M. C. (2005). A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2):117–132.

Dingli, D. and Michor, F. (2006). Successful Therapy Must Eradicate Cancer Stem Cells. *Stem Cells*, 24(12):2603–2610.

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P., Grönberg, H., Hultman, C. M., and

McCarroll, S. A. (2014). Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*, 371(26):2477–2487.

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Ohshima, K., Uchida, N., Taniguchi, S., Ohara, O., Akashi, K., Harada, M., and Shultz, L. D. (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nature Biotechnology*, 25(11):1315.

Jaiswal, S. and Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science*, 366(6465).

Komarova, N. L. and Wodarz, D. (2007). Effect of cellular quiescence on the success of targeted CML therapy. *PLoS ONE*, 2(10).

Kumar, S. and Geiger, H. (2017). HSC Niche Biology and HSC Expansion Ex Vivo. *Trends in Molecular Medicine*, 23(9):799–819.

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E., O'Neill, L., Stratton, M. R., Laurenti, E., Green, A. R., Kent, D. G., and Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*.

Meiss, J. D. (2007). *Differential Dynamical Systems*. Society for Industrial and Applied Mathematics.

Nie, Y., Han, Y.-C., and Zou, Y.-R. (2008). CXCR4 is required for the quiescence of primitive hematopoietic cells. *Journal of Experimental Medicine*, 205(4):777–783.

Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertoncello, I., Bendall, L. J., Simmons, P. J., and Haylock, D. N. (2005). Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood*, 106(4):1232–1239.

Park, D. S., Akuffo, A. A., Muench, D. E., Grimes, H. L., Epling-Burnette, P. K., Maini, P. K., Anderson, A. R. A., and Bonsall, M. B. (2019). Clonal hematopoiesis of indeterminate potential and its impact on patient trajectories after stem cell transplantation. *PLOS Computational Biology*, 15(4).

Pedersen, R. K., Andersen, M., Knudsen, T. A., Sajid, Z., Gudmand-Hoeyer, J., Dam, M. J. B., Skov, V., Kjær, L., Ellervik, C., Larsen, T. S., Hansen, D., Pallisgaard, N., Hasselbalch, H. C., and Ottesen, J. T. (2020). Data-driven analysis of JAK2 V617F kinetics during interferon-alpha2 treatment of patients with polycythemia vera and related neoplasms. *Cancer Medicine*, 9(6):2039–2051.

Pedersen, R. K., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., Ottesen, J. T., and Stiehl, T. (n.d.). Combining stem cell mobilization with preconditioning: Insights about competition in the stem cell niche from mathematical modelling. *Submitted*.

Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859):105–111.

Roeder, I. and Loeffler, M. (2002). A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Experimental Hematology*, 30(8):853–861.

Stiehl, T., Baran, N., Ho, A. D., and Marciniak-Czochra, A. (2015). Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: A model to predict patient survival. *Cancer Research*, 75(6):940–949.

Stiehl, T., Wang, W., Lutz, C., and Marciniak-Czochra, A. (2020). Mathematical modeling provides evidence for niche competition in human AML and serves as a tool to improve risk stratification. *Cancer Research*.

Takahashi, K., Patel, K. P., Kantarjian, H., Luthra, R., Pierce, S., Cortes, J., and Verstovsek, S. (2013). JAK2 p.V617F detection and allele burden measurement in peripheral blood and bone marrow aspirates in patients with myeloproliferative neoplasms. *Blood*, 122(23):3784–3786.

Vaidya, A. and Kale, V. (2015). Hematopoietic Stem Cells, Their Niche, and the Concept of Co-Culture Systems: A Critical Review. *Journal of stem cells*, 10(1):13–31.

Wang, W., Stiehl, T., Raffel, S., Hoang, V. T., Hoffmann, I., Poisa-Beiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Bochtler, T., Wuchter, P., Essers, M., Jauch, A., Trumpp, A., Marciniak-Czochra, A., Ho, A. D., and Lutz, C. (2017). Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia. *Haematologica*, 102(9):1567–1577.

Wilson, A. and Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nature Reviews Immunology*, 6(2):93–106.

Yang, S.-H., Wang, T.-F., Tsai, H.-H., Lin, T.-Y., Wen, S.-H., and Chen, S.-H. (2009). Preharvest hematopoietic progenitor cell counts predict CD34+ cell yields in granulocyte-colony-stimulating factor-mobilized peripheral blood stem cell harvest in healthy donors. *Transfusion*, 50(5):1088–1095.

Zhang, Y. and Gao, Y. (2016). Novel chemical attempts at ex vivo hematopoietic stem cell expansion. *International Journal of Hematology*, 103(5):519–529.

Supplementary material

A Arbitrary number of division before HSC exhaustion

In general, the number of divisions HSC perform before entering the state of exhaustion can be larger than one. In the following, we present a model where HSC become exhausted after an arbitrary finite number of J divisions. Under certain assumptions, this model is identical to the model derived in the main text. For the sake of simplicity we consider a version of the model describing only a single clone. We subdivide the compartment of active cells A to keep track of the number of divisions cells have performed. By A_i we denote the population of activated HSC that have performed $i-1$ divisions. Hence, cells that have just detached from the niche belong to compartment A_1, upon division each cell of compartment A_1 gives rise to two cells of compartment A_2, et cetera.

The model with J HSC divisions before exhaustion can be written as:

$$\dot{N} = \sum_{i=1}^{J} (b_i(K-N)A_i) - uN \tag{A.1a}$$

$$\dot{A}_1 = -b_1(K-N)A_1 - (r_1+d_1)A_1 + uN \tag{A.1b}$$

$$\dot{A}_i = -b_i(K-N)A_i + 2r_{i-1}A_{i-1} - (r_i+d_i)A_i \tag{A.1c}$$

$$\dot{A}_J = -b_J(K-N)A_J + 2r_{J-1}A_{J-1} - d_JA_J \tag{A.1d}$$

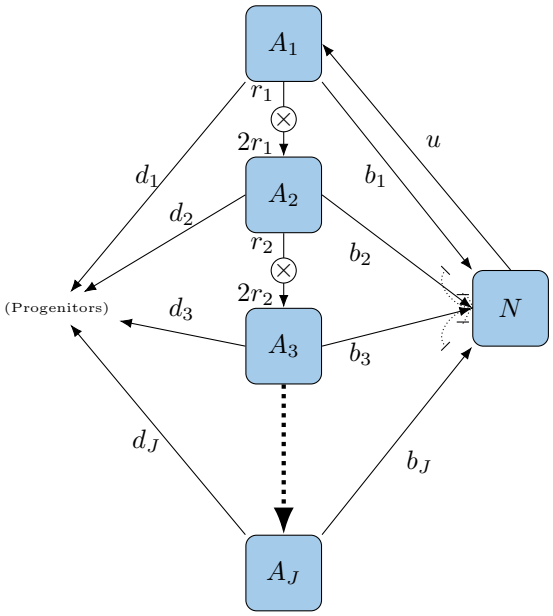


Figure A1: **Compartment diagram of the model with multiple stage of free HSC.** Each box illustrates a variable, with arrows depicting the relations between the variables. The circles with the crosses, $\times$ signify a doubling.

By a quasi-steady-state approximation of A_2 type cells, we find that

$$\dot{A}_2 = 0 \Rightarrow A_2 = \frac{2r_1}{r_2+d_2+b_2(K-N)}A_1 \tag{A.2}$$

The expression for $\dot{A}_3$ can then be written as:

$$\dot{A}_3 = 2r_2\left(\frac{2r_1}{r_2+d_2+b_2(K-N)}\right)A_1 - d_3A_3 - b_3(K-N)A_3 \tag{A.3}$$

For the model presented in our previous work (Pedersen et al., nd), fitting to data revealed that b_{A_1} is very low. Here this corresponds to b_1. Assuming that this also holds for the intermediate steps, we set $b_2 = 0$. Equation (A.3) then becomes

$$\dot{A}_3 = 4r_1\rho_2A_1 - d_3A_3 - b_3(K-N)A_3 \tag{A.4}$$

where $\rho_2 = \dfrac{r_2}{r_2 + d_2}$.

Repeating this for all division events taking place before exhaustion, i.e. assuming $b_i = 0$ and making a quasi-steady-state approximation of step i, the final step, A_J, becomes:

$$\dot{A_J} = 2r_1 \left(\prod_{i=2}^{J} 2\rho_i \right) A_1 - d_J A_J - b_J (K - N) A_J \tag{A.5}$$

Defining $\gamma = \prod_{i=2}^{J} 2\rho_i$ we obtain the system described in the main text. In addition, if the probability r_i that a stem cell in compartment A_i divides is approximately equal to the probability that it differentiates, d_i, we can approximate $2\rho_i = \frac{2r_i}{r_i + d_i} \approx 1$. This implies that $\gamma = 1$ which further simplifies the system. Hence the final step, A_J, is defined as presented in the main paper.

B Details about stability of steady states

In section 3.3 of the main text, we consider the stability of the steady states for the full $3n$ system.

For the trivial steady state S_0^* and the single-clone steady states S_j^*, this depends solely on the eigenvalues of the matrix D_j defined as:

$$D_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j - u_j & b_{I_j} \left(K - \sum_{i=1}^{n} N_i\right) & b_{A_j} \left(K - \sum_{i=1}^{n} N_i\right) \\ b_{I_j} I_j & -b_{I_j} \left(K - \sum_{i=1}^{n} N_i\right) - d_{I_j} & 2\gamma r_j \\ u_j + b_{A_j} A_j & 0 & -b_{A_j} \left(K - \sum_{i=1}^{n} N_i\right) - r_j - d_{A_j} \end{pmatrix} \tag{B.1}$$

The characteristic polynomial of D_j, can be written as $P_j(\lambda) = \lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$ for which:

$$\begin{aligned} a_2 =& (b_{I_j} + b_{A_j}) N_E + b_{I_j} I_j + b_{A_j} A_j + d_{I_j} + d_{A_j} + r_j + u_j \\ a_1 =& b_{A_j} b_{I_j} N_E^2 + b_{I_j} (b_{A_j} A_j + b_{A_j} I_j + b_{A_j} \frac{d_{I_j}}{b_{I_j}} + d_{A_j} + r_j + u_j) N_E \\ & + (d_{A_j} + d_{I_j} + r_j) \left(b_{A_j} A_j + b_{I_j} I_j + u_j\right) + d_{I_j} (r_j + d_{A_j}) \\ a_0 =& b_{I_j} \left[b_{A_j} d_{I_j} I_j - ((2\gamma - 1) r_j - d_{A_j}) b_{A_j} A_j - ((2\gamma - 1) r_j - d_{A_j}) u_j \right] N_E \\ & + d_{I_j} (r_j + d_{A_j}) \left(b_{A_j} A_j + b_{I_j} I_J + u_j\right) \end{aligned}$$

where $N_E = K - \sum_{i=1}^{n} N_i$

The Routh-Hurwitz criterion can be used to determine if all roots of the polynomial P_j have negative real part. This is the case if and only if $a_2 > 0$, $a_1 > 0$, $a_0 > 0$ and $a_2 a_1 > a_0$.

It can be seen from the above expressions that when N_E, A_j and I_j are non-negative, it must hold that $a_2 > 0$ as well as $a_1 > 0$. Hence when D_j is evaluated in either of the feasible steady states, it is the case that $a_2 > 0$ and $a_1 > 0$.

Whenever N_E, A_j and I_j are non-negative, we find that $a_2 a_1 - a_0 > 0$. This can be seen by considering the negative parts of $-a_0$. Each negative contribution also appears in $a_2 a_1$ and hence only positive contributions of $-a_0$ and $a_2 a_1$ remain, and thus $a_2 a_1 - a_0 > 0$

Thus, the Routh-Hurwitz stability criterion depends only on the sign of a_0.

When evaluating in the trivial steady state, S_0^*, we find that the characteristic polynomial of $D_j|_{S_0^*}$ has $a_0|_{S_0^*} = -u_j b_{I_j} \left((2\gamma - 1) r_j - d_{A_j}\right) \left(K - \frac{1}{F_j}\right)$. The requirement for feasibility of a single-clone steady state S_j^* implies that $(2\gamma - 1) r_j - d_{A_j} > 0$ as well as $K - \frac{1}{F_j} > 0$, and hence if the j'th single-clone steady state is feasible, $a_0|_{S_0^*} < 0$, and hence the Routh-Hurwitz stability criterion does not hold and $D_j|_{S_0^*}$ has at least one eigenvalue with positive real part.

Evaluating in the single-clone steady state, S_j^*, we find $a_0|_{S_j^*} = -a_0|_{S_0^*}$, and the Routh-Hurwitz stability criterion holds. Hence, all eigenvalues of $D_j|_{S_j^*}$ have negative real part.

Finally we consider $D_j|_{S_k^*}$ where $j \neq k$, i.e. the matrix related to the j'th clone evaluated at the k'th single-clone steady state. Again, assuming the j'th clone has feasible steady states, only the sign of a_0 has to be considered.

To determine the eigenvalues of $D_j|_{S_k^*}$ for $j \neq k$, we find that the constant term of the characteristic polynomial is:

$$a_{0, D_j|_{S_k^*}} = u_j b_{I_j} \left((2\gamma - 1) r_j - d_{A_j}\right) \frac{1}{F_k} \left(\frac{F_k}{F_j} - 1\right) \tag{B.2}$$

If $F_k > F_j$ we find $a_{0,D_j|_{S_k^*}} > 0$, and the Routh-Hurwitz stability criterion holds, and the eigenvalues of $D_j|_{S_k^*}$ all have negative real part. Conversely, when $F_k < F_j$, the criterion does not hold, and at least one eigenvalue has positive real part.

Co-existence steady state

When two or more cell-type have equal fitness, a simplex of steady states exists, S_C^*. We here show that the system has an eigenvector along the simplex, and that the corresponding eigenvalue is zero. For simplicity, we show this for the case where only two cell-types are considered.

In this particular case, the Jacobian is given as:

$$Jac = \begin{pmatrix} D_1 & G_1 \\ G_2 & D_2 \end{pmatrix} \tag{B.3}$$

When projecting the simplex of steady states onto a N_1, N_2 plane, it is a line with slope -1. The particular non-zero eigenvector corresponding to the eigenvalue zero can be written as:

$$V_0 = \begin{pmatrix} 1 \\ \nu_{I_1} \\ \nu_{A_1} \\ -1 \\ -\nu_{I_2} \\ -\nu_{A_2} \end{pmatrix} \tag{B.4}$$

where ν_{A_j} and ν_{I_j} are defined as

$$\nu_{A_j} = \frac{u_j}{b_{A_j}\frac{1}{F_j} + r_j + d_{A_j}} \tag{B.5}$$

and

$$\nu_{I_j} = \frac{2\gamma r_j u_j}{\left(r_j + d_{A_j} + b_{A_j}\frac{1}{F_j}\right)\left(d_{I_j} + b_{I_j}\frac{1}{F_j}\right)} = \frac{2\gamma r_j \nu_{A_j}}{d_{I_j} + b_{I_j}\frac{1}{F_j}} \tag{B.6}$$

Note that the steady state values of A_j and I_j can be written in general as:

$$A_j^* = \nu_{A_j} N_j^* \tag{B.7}$$

and

$$I_j^* = \nu_{I_j} N_j^* \tag{B.8}$$

The Jacobian multiplied by the eigenvector V_0 yields:

$$JacV_0 = \begin{pmatrix} -u_1 + \nu_{I_1} b_{I_1} N_E + \nu_{A_1} b_{A_1} N_E \\ -\nu_{I_1}(b_{I_1} N_E + d_{I_1}) + \nu_{A_1} 2\gamma r_1 \\ u_1 - \nu_{A_1}(b_{A_1} N_E + r_1 + d_{A_1}) \\ -u_2 + \nu_{I_2} b_{I_2} N_E + \nu_{A_2} b_{A_2} N_E \\ -\nu_{I_2}(b_{I_2} N_E + d_{I_2}) + \nu_{A_2} 2\gamma r_2 \\ u_2 - \nu_{A_2}(b_{A_2} N_E + r_2 + d_{A_2}) \end{pmatrix} \tag{B.9}$$

For existence of the co-existence steady state, we have $F = F_1 = F_2$. Evaluating at S_C^* we find:

$$Jac|_{S_C^*} V_0 = \begin{pmatrix} -u_1 + \nu_{I_1} b_{I_1} F^{-1} + \nu_{A_1} b_{A_1} F^{-1} \\ -\nu_{I_1}(b_{I_1} F^{-1} + d_{I_1}) + \nu_{A_1} 2\gamma r_1 \\ u_1 - \nu_{A_1}(b_{A_1} F^{-1} + r_1 + d_{A_1}) \\ -u_2 + \nu_{I_2} b_{I_2} F^{-1} + \nu_{A_2} b_{A_2} F^{-1} \\ -\nu_{I_2}(b_{I_2} F^{-1} + d_{I_2}) + \nu_{A_2} 2\gamma r_2 \\ u_2 - \nu_{A_2}(b_{A_2} F^{-1} + r_2 + d_{A_2}) \end{pmatrix} \tag{B.10}$$

All terms of this vector are zero. This can be seen by substituting F with $F_1 = \frac{b_{I_1}((2\gamma-1)r_1 - d_{A_1})}{d_{I_1}(r_1 + d_{A_1})}$ in the first three components and with $F_2 = \frac{b_{I_2}((2\gamma-1)r_2 - d_{A_2})}{d_{I_2}(r_2 + d_{A_2})}$ in the bottom three components. Thus, we find $Jac|_{S_C^*} V_0 = 0V_0$, and the vector V_0 is indeed a non-zero eigenvector of $Jac|_{S_C^*}$, with the corresponding eigenvalue of 0.

C Details about analysis of the transformed model

In section 4.2 we present a transformation of the reduced model. When only two distinct clones are considered, the 2-*clone transformed model* is given as:

$$\dot{T} = (g_1(T)(1 - C) + g_2(T)C)\, T \tag{C.1a}$$

$$\dot{C} = (g_2(T) - g_1(T))(1 - C)C \tag{C.1b}$$

where $g_j = u_j \alpha_j f_j \frac{1 - T - f_j^{-1}}{\alpha_j + 1 - T}$. Only solutions to equations (C.1) that exists on the domain $\mathbb{D} = \{(T, C) \in (0, 1] \times [0, 1]\}$ are considered. In this section, we present a detailed analysis of equations (C.1).

Initially, note that solutions are indeed restricted to the domain $\mathbb{D}$. This can be observed by considering equations (C.1) along the boundaries of the domain. We first consider the lower boundary of T by taking $T \to 0^+$:

$$\lim_{T \to 0^+} \frac{\dot{T}}{T} = g_1(0)(1 - C) + g_2(0)C \tag{C.2}$$

In general, we assume that feasible steady states exists for both clones, and hence $0 < f_1^{-1} < 1$ and $0 < f_2^{-1} < 1$. This implies that $g_1(0) > 0$ and $g_2(0) > 0$, and hence $\lim_{T \to 0^+} \frac{\dot{T}}{T} > 0$, and trajectories of solutions with initial conditions within $\mathbb{D}$ do not approach $T = 0$. The restrictions of the remaining three boundaries are determined by evaluating the differential equations at the boundaries:

$$\dot{T}|_{T=1} = g_1(1)(1 - C) + g_2(1)C = -u_1(1 - C) - u_2 C < 0 \tag{C.3}$$

$$\dot{C}|_{C=0} = 0 \tag{C.4}$$

$$\dot{C}|_{C=1} = 0 \tag{C.5}$$

Equilibria

While the steady states for the n-clone transformed model were presented in the main text, we briefly re-iterate the steady states of the 2-clone form. $\dot{C} = 0$ implies either $C = 0$, $C = 1$ or $g_2(T) = g_1(T)$. We consider these in order. For $C = 0$, $\dot{T} = g_1(T)T$, which has just one zero, namely $T = T_1^* = 1 - f_1^{-1}$. This results in the equilibrium $\overline{S_1} = (1 - f_1^{-1}, 0)$. Similarly, $C = 1$ implies $\dot{T} = g_2(T)T$ which has zero $T = T_2^* = 1 - f_2^{-1}$, resulting in the equilibrium $\overline{S_2} = (1 - f_2^{-1}, 1)$.

Lastly we consider co-existence steady states, $\overline{S_C}$, where $0 < C_c^* < 1$. We note that $\dot{T}$ can be re-written such that:

$$\dot{T} = (g_1(T) + (g_2(T) - g_1(T))\, C)\, T \tag{C.6}$$

When $g_2(T) = g_1(T)$, we can write $\dot{T} = g_1(T)T$, which is zero only if $T = T_1^*$. Hence, an equilibrium with $0 < C_c^* < 1$ must have $T = T_1^*$. Since $g_j(T) = 0$ only holds for $g_j(1 - f_j^{-1})$, $\dot{T} = 0$ implies $g_2(T_1^*) = 0$, and hence $f_1 = f_2$. Thus, co-existence steady states $\overline{S_C}$ only exist when $f_1 = f_2$. For $f_1 \neq f_2$, equations (C.1) has just two equilibria, $\overline{S_1}$ and $\overline{S_2}$.

Lemma C.1: Equilibria of the 2-clone transformed model

For $f_1 \neq f_2$, equations (C.1) has just two equilibria, $\overline{S_1}$ and $\overline{S_2}$. In order (T, C) these are:

$$\overline{S_1} = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix} \quad \text{and} \quad \overline{S_2} = \begin{pmatrix} 1 - f_2^{-1} \\ 1 \end{pmatrix} \tag{C.7}$$

When $f_1 = f_2$, a line of steady states $\overline{S_C}$ exist:

$$\overline{S_C} = \begin{pmatrix} 1 - f_1^{-1} \\ \nu \end{pmatrix} \tag{C.8}$$

where $\nu \in [0, 1]$.

For $f_1 \neq f_2$, the two equilibria $\overline{S_1}$ and $\overline{S_2}$ exist on the boundary of $\mathbb{D}$. Hence an orbit of solutions cannot enclose either of the equilibria. For $f = f_1 = f_2$, the line of steady states $\overline{S_C}$ divides $\mathbb{D}$ in two sub-domains, one with $T \in (0, 1 - f^{-1})$ and one with $T \in (1 - f^{-1}, 1]$. Trajectories in one sub-domain cannot enter the other as it would require leaving the domain $\mathbb{D}$ or crossing $\overline{S_C}$, neither of which is possible. Hence, periodic orbits surrounding $\overline{S_C}$ cannot exist. As any periodic orbit in $2D$ must have at least one interior equilibrium due to index theory (Meiss, 2007), there can be no periodic solutions to equations (C.1). Thus, due to the *Poincaré-Bendixon* theorem, any solution trajectory must converge to either $\overline{S_1}$ or $\overline{S_2}$ when $f_1 \neq f_2$, or to a point on $\overline{S_C}$ for $f_1 = f_2$.

Global attraction of steady states

It can be shown trivially that when $f_1 = f_2$, $\dot{T} > 0$ for all $T \in (0, 1 - f^{-1})$ and $\dot{T} < 0$ for $T \in (1 - f^{-1}, 1]$. Hence the points on the line $\overline{S_C}$ are attractive for all initial conditions. Note however, as we show below, C is not necessarily constant when $f_1 = f_2$, see figure A2.

For $f_1 \neq f_2$, an attractive trapping region can be found. For brevity, we consider the case $f_1 < f_2$, as the case $f_1 > f_2$ follows by symmetry.

Defining

$$T_{2,\epsilon} = 1 - f_2^{-1} + \epsilon \tag{C.9}$$

for any $\epsilon \geq 0$, we observe that $g_1(T_{2,\epsilon}) < 0$ and $g_2(T_{2,\epsilon}) \leq 0$, with equality only for $\epsilon = 0$. Hence

$$\dot{T}|_{T_{2,\epsilon}} = (g_1(T_{2,\epsilon})(1 - C) + g_2(T_{2,\epsilon})C)\, T_{2,\epsilon} \geq 0 \tag{C.10}$$

where $\dot{T}|_{T_{2,\epsilon}} = 0$ only when both $C = 1$ and $\epsilon = 0$.

We present an argument analogous to our definition of the trapping region of the full model in section 3.1 of the main text.

Consider an interval $\mathbb{T} = [T_{2,\epsilon}, T_{2,\epsilon} + \Delta]$ with $\Delta > 0$ on which $\dot{T} < 0$. Due to the extreme value theorem, there exists a finite η such that $-\dot{T} \geq \eta, \forall T \in \mathbb{T}$, since $\mathbb{T}$ is a compact subset. In particular

$$\inf_{\mathbb{T}}(-\dot{T}) = \min_{\mathbb{T}}(-\dot{T}) = \eta > 0 \tag{C.11}$$

and hence $\frac{1}{-\dot{T}} \leq \frac{1}{\eta}$ holds for all $T \in \mathbb{T}$.

To determine the time t_Δ it takes for the flow from $T(0) = T_{2,\epsilon} + \Delta$ to reach $T(t_\Delta) = T_{2,\epsilon}$ we integrate over the interval:

$$t_\Delta = \int_0^{t_\Delta} dt = \int_{T(0)}^{T(t_\Delta)} \frac{1}{\dot{T}} dT = \int_{T_{2,\epsilon}}^{T_{2,\epsilon}+\Delta} \frac{1}{-\dot{T}} dT \leq \frac{\Delta}{\eta} \tag{C.12}$$

and hence t_Δ is finite. Thus for any $\Delta > 0$, any point with $T(0) = T_{2,\epsilon} + \Delta$ will move along a trajectory that enters the region $T \leq T_{2,\epsilon}$ in finite time for any $\epsilon > 0$.

Defining

$$T_{1,-\epsilon} = 1 - f_1^{-1} - \epsilon \tag{C.13}$$

a similar argument can be made that any point with $T(0) < T_{1,-\epsilon}$ enter the region in finite time. Hence there is a region $\mathbb{M} = \{(T, C) \in [1 - f_1^{-1}, 1 - f_2^{-1}] \times (0, 1)\} \subset \mathbb{D}$ which is an attractive trapping region for all points in $\mathbb{D} \setminus \mathbb{M}$

Observe that for all $T \in \mathbb{M}$, it holds that $g_1(T) \leq 0$ with equality only in $T = T_1^*$, and $g_2(T) \geq 0$ with equality only in $T = T_2^*$. Hence $g_2(T) - g_1(T) > 0$ since $T_1^* \neq T_2^*$. Thus, for $T \in \mathbb{M}$:

$$\lim_{C\to 0^+} \frac{\dot{C}}{C} = \lim_{C\to 0^+} \left(g_2(T) - g_1(T)\right)(1-C) = g_2(T) - g_1(T) > 0 \tag{C.14}$$

and hence $\overline{S_1}$ is unstable.

We define $\mathbb{M}_\epsilon = \{(T, C) \in [1 - f_1^{-1}, 1 - f_2^{-1}] \times (\epsilon, 1)\}$ for any $\epsilon > 0$. The region $\mathbb{M}_\epsilon$, only contains one equilibrium, $\overline{S_2}$. As there can be no periodic solutions, we have determined that $\overline{S_2}$ is attractive for all trajectories with $C(0) > 0$. As $\mathbb{M}$ is an attractive trapping region for all points in $\mathbb{D} \setminus \mathbb{M}$, this result extends to all points in $\mathbb{D}$. For the case $f_1 > f_2$, a similar construction with $C \in [0, 1 - \epsilon]$ for all $\epsilon > 0$ can be made, implying that $\overline{S_1}$ is an attractor for all initial conditions with $C(0) < 1$.

Lemma C.2: Limit sets of solutions

For $f_1 < f_2$, there are two equilibria, $\overline{S_1}$ and $\overline{S_2}$. $\overline{S_2}$ is an attractor for all points with $C(0) > 0$.
For $f_1 > f_2$, there are two equilibria, $\overline{S_1}$ and $\overline{S_2}$. $\overline{S_1}$ is an attractor for all points with $C(0) < 1$.

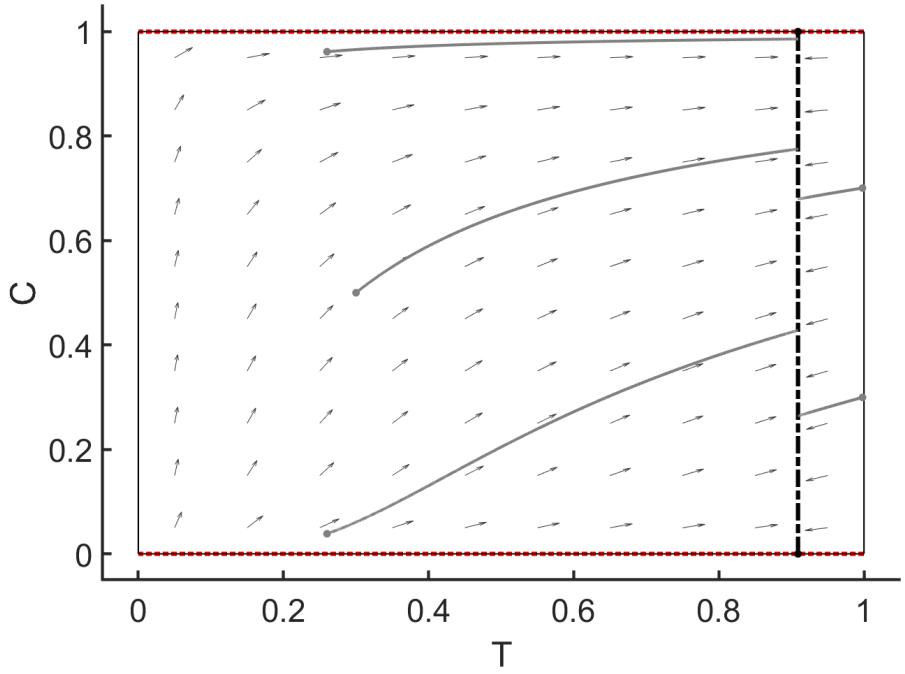


Figure A2: **Illustrative example of an equal fitness scenario.** In the simulation, $\rho_1 = \rho_2$ and $\alpha_1 = \alpha_2$ and hence $f_1 = f_2$. However $u_2 = 5u_1$, which causes the abundance of the clone with the higher fitness to increase as solutions approach the co-existence equilibrium.

Conditions for $g_2(T) = g_1(T)$.
The expression $g_2(T) = g_1(T)$ can be written as a second order polynomial in T. Defining $\mu_j = u_j \alpha_j f_j$:

$$g_2(T) = g_1(T) \tag{C.15}$$

$$u_2\alpha_2 f_2 \frac{1 - T - f_2^{-1}}{\alpha_2 + 1 - T} = u_1\alpha_1 f_1 \frac{1 - T - f_1^{-1}}{\alpha_1 + 1 - T} \tag{C.16}$$

$$\mu_2 \frac{1 - T - f_2^{-1}}{\alpha_2 + 1 - T} = \mu_1 \frac{1 - T - f_1^{-1}}{\alpha_1 + 1 - T} \tag{C.17}$$

$$aT^2 + bT + c = 0 \tag{C.18}$$

where

$$a = \mu_2 - \mu_1 \tag{C.19a}$$
$$b = -\mu_2\left((1 - f_2^{-1}) + (\alpha_1 + 1)\right) + \mu_1\left((1 - f_1^{-1}) + (\alpha_2 + 1)\right) \tag{C.19b}$$
$$c = \mu_2(1 - f_2^{-1})(\alpha_1 + 1) - \mu_1(1 - f_1^{-1})(\alpha_2 + 1) \tag{C.19c}$$

If $a = 0$, $b = 0$ and $c = 0$, then g_1 and g_2 are identical functions. Assuming $\mu_1 = \mu_2 = \mu$, $a = 0$ is fulfilled. Equation (C.19b) then becomes:

$$b = \mu\left(f_2^{-1} - f_1^{-1} + \alpha_2 - \alpha_1\right) \tag{C.20}$$

For $b = 0$, it must hold that $f_1^{-1} = f_2^{-1} + \alpha_2 - \alpha_1$. This allows for a simplification of c:

$$c = \mu\left(f_2^{-1} + \alpha_1\right)(\alpha_2 - \alpha_1) \tag{C.21}$$

And hence $\alpha_2 = \alpha_1$ implies $c = 0$, since all parameters are positive.

Writing up $g_2(T) = g_1(T)$ again, with $\alpha = \alpha_1 = \alpha_2$, we obtain:

$$\mu_2\frac{1 - T - f_2^{-1}}{\alpha + 1 - T} = \mu_1\frac{1 - T - f_1^{-1}}{\alpha + 1 - T} \tag{C.22}$$
$$\mu_2(1 - T - f_2^{-1}) = \mu_1(1 - T - f_1^{-1}) \tag{C.23}$$
$$\mu_2(1 - T - f_2^{-1}) - \mu_1(1 - T - f_1^{-1}) = 0 \tag{C.24}$$
$$-(\mu_2 - \mu_1)T + \mu_2(1 - f_2^{-1}) - \mu_1(1 - f_1^{-2}) = 0 \tag{C.25}$$

which is linear in T. If $\mu_1 = \mu_2$, the expression simplifies to $f_1 = f_2$.

Finally, when both $f_1 = f_2$ and $\alpha_1 = \alpha_2$, the functions g_1 and g_2 are identical only when $u_1 = u_2$. Thus, it is only possible for $g_2(T) = g_1(T)$ to hold for all T when all parameters are equal.

Lemma C.3: Condition for $g_2 = g_1$

The functions $g_1(T)$ and $g_2(T)$ are equal for all T if and only if $f_1 = f_2$, $\alpha_1 = \alpha_2$ and $u_1 = u_2$, in which case g_1 and g_2 are identical functions.

We emphasize that $f_1 = f_2$ is *not* sufficient for $g_2(T) = g_1(T)$ to hold for all T.

Nullclines for $\dot{T}$ and $\dot{C}$

Assuming that g_1 and g_2 are not identical, and hence that $f_1 \neq f_2$, $\alpha_1 \neq \alpha_2$ and $u_1 \neq u_2$, it is possible for the second order polynomial of equation (C.18) to have one or two real roots between 0 and 1. These values, $T_\pm$, are nullclines for $\dot{C}$, shown in red in the figures of this section and in the main text. It is possible for both $T_\pm$ to be between 0 and 1 simultaneously, see figure A3, and for neither to be between 0 and 1.

For equal fitness, $f = f_1 = f_2$, the differential equation for T can be written:

$$\dot{T} = \left(\frac{u_1\alpha_1}{\alpha_1 + 1 - T}(1 - C) + \frac{u_2\alpha_2}{\alpha_2 + 1 - T}C\right)(1 - T - f^{-1})fT \tag{C.26}$$

$\dot{T} = 0$ has two solutions: $T_f^* = 1 - f^{-1}$ and

$$T_o^* = 1 + \alpha_1\alpha_2\frac{u_1(1 - C) + u_2C}{u_1\alpha_1(1 - C) + u_2\alpha_2C} \tag{C.27}$$

Since $C \in [0, 1]$, and all parameters are positive, $T_o^* > 1$ and hence not feasible. Thus, in a scenario with equal fitness, the sign of $\dot{T}$ only changes at $T_f^* = 1 - f^{-1}$.

For $f_1 \neq f_2$, the sign of $\dot{T}$ is determined by $g_1(T)(1 - C) + g_2(T)(C)$, from which a function $C_{null}(T)$ can be defined:

$$C_{null}(T) = \frac{g_1(T)}{g_1(T) - g_2(T)} \tag{C.28}$$

$\dot{T} = 0$ for all $(T, C) = (T_n, C_{null}(T_n))$ and hence $C_{null}(T)$ is a nullcline for T. In the two single-clone steady states, we find $C_{null}(1 - f_1^{-1}) = 0$ and $C_{null}(1 - f_2^{-1}) = 1$.

Observe that

$$C_{null}(T) \underset{g_1(T)\to g_2(T)}{\longrightarrow} \pm\infty \tag{C.29}$$

and thus the $C_{null}(T)$ is undefined for $T = T_\pm$ and the nullclines do not cross, as expected from the analysis of steady states above.

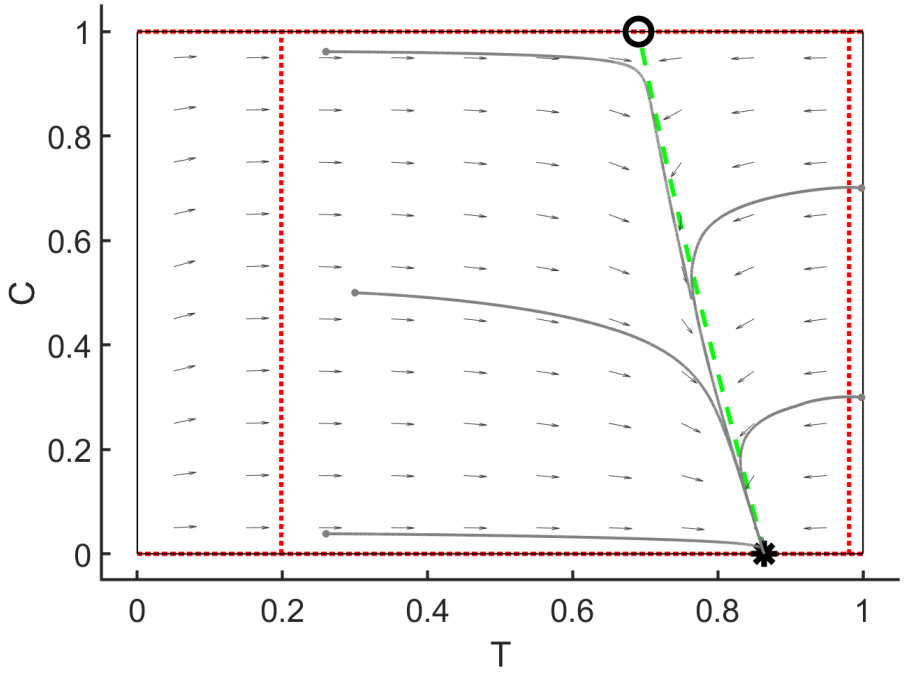


Figure A3: **Phase plane illustration of two $\dot{C}$-nullclines.** The figure illustrates a particular choice of parameters such that the two roots of $g_2(T) - g(1) = 0$, $T_\pm$, are both between 0 and 1. The particular choice of parameter for the figure were $\rho_1 = 0.529$, $\rho_2 = 0.627$, $\alpha_1 = 0.008$, $\alpha_2 = 0.079$, $u_1 = 0.113$ and $u_2 = 0.038$.