Long-term clonal analysis using stochastic models reveals heterogeneity and quiescence of hematopoietic stem cells

Yuri Garcia Vilela [a, b], Lars Thielecke [b], Artur C. Fassoni [b, c], Ingmar Glauche [b]

[a] Department of Mathematics, Universidade Federal de Minas Gerais, Av. Pres. Antônio Carlos, 6627, 31270-901, Belo Horizonte, MG, Brazil.
[b] Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Fetscherstr. 74, D-01307 Dresden, Germany.
[c] Institute of Mathematics and Computing, Universidade Federal de Itajubá, Av. B P S, 1303, 37500-185, Itajubá, MG, Brazil.

Authors:
Yuri Garcia Vilela <ygvilela@ufmg.br>
Lars Thielecke <lars.thielecke@tu-dresden.de>
Artur César Fassoni <fassoni@unifei.edu.br>

Corresponding author:
Ingmar Glauche <ingmar.glauche@tu-dresden.de>

Abstract

Hematopoietic stem cells (HSCs) maintain lifelong production of blood by balancing self-renewal and differentiation. However, certain aspects of their divisional dynamics, namely the role of quiescence and the intrinsic heterogeneity of the HSC pool, are not completely understood. High-resolution clonal tracking provides a powerful resource to investigate such dynamics as the data captures patterns of clonal persistence, dilution and late clonal emergence. Here, we apply mechanistic mathematical modeling to longitudinal clonal data from non-human primates to explore structural requirements that underlie the observed dynamical patterns.

We show that models treating HSCs as a single, homogeneous population can explain the gradual loss of clonal diversity, but fail to reproduce clone size distributions and the long-term persistence of small and late-appearing clones. To address this, we propose a stochastic, two-compartment model in which HSCs transition reversibly between an actively cycling state and a quiescent, potentially niche-bound state. Compared to the simpler one-compartment model, this advanced framework provides a substantially improved fit for different metrics, consistently captures clone size distributions and explains the delayed activation and sustained coexistence of small and large clones.

These results provide quantitative evidence that heterogeneity within the HSC pool, particularly the existence of a reversible quiescent state, is critical to account for clonal aspects of long-term hematopoiesis. Our findings highlight how clonal data can uncover underlying regulatory mechanisms and supports a central role for niche-mediated HSC quiescence in maintaining stable and diverse blood production over time.

Keywords: Hematopoietic stem cells, quiescence, clonal tracking, stochastic modeling

Highlights:

- Clonal data offers a robust approach to explore hematopoietic stem cell dynamics
- Homogeneous HSC population cannot reproduce observed long-term clonal patterns
- Niche-mediated quiescence improves model fit across different clonality metrics
- Model results suggest reversible transition between HSC quiescence and activation

Introduction

Hematopoietic stem cells (HSCs) are crucial for sustained blood cell production throughout life by balancing self-renewal and differentiation. These rare, multipotent cells reside predominantly in the bone marrow and generate all blood lineages while at the same time maintaining their own stem cell pool. HSC activity is regulated by both intrinsic genetic and epigenetic mechanisms as well as external signals from the bone marrow microenvironment, also referred to as the stem cell niche [8, 9, 31, 29]. Few studies addressed the dynamical regulation of stem cell numbers and divisional activity with a particular focus on how the niche can regulate the balance between self-renewal and differentiation and thereby influence stem cell numbers [41, 13, 14, 27].

Clonal studies of HSC contributions rely on the identification and tracking of the direct offspring of individual stem cells, which can be achieved through artificial marking events, either in vitro [12, 30, 23, 6] or in vivo [28, 38], or by tracing naturally occurring mutations [21, 24]. By monitoring individual clones over time, such studies have elucidated key aspects of HSC self-renewal, long-term persistence and aging, shaping our understanding of hematopoiesis and its regulatory mechanisms [38, 23, 39, 24]. Clonal analysis also elucidated how HSCs and other progenitor cells contribute to blood production and how these differentiation pathways are challenged under stress conditions [3]. While gene therapy settings are a possible access to study post-transplantation clonality in humans, many aspects of dynamical engraftment and long-term polyclonality are still incompletely understood.

Non-human primates (NHPs) serve as an important model for studying clonal dynamics of HSCs in a more clinically relevant context [20, 32]. A remarkable contribution was the recent work by Radtke et al. [33] that followed individual HSC and progenitor clones over extended periods after autologous transplantation. The high-resolution tracking revealed that > 90% of long-term persisting HSCs contributed already to early hematopoietic recovery, challenging the assumption that only committed progenitors drive short-term hematopoiesis. Additionally, while their data showed a rapid loss of unique clones and the formation and expansion of long-term persisting clones, it also revealed the continued detection of new clones over time.

Radtke and colleagues use a mathematical, one-compartment model of HSC competition to quantitatively explain the manifestation of long-term persisting clones. This feature is known as clonal conversion and represents a hallmark of neutral competition in which stochastic processes lead to the extinction of some clones and the growth of others. The underlying model assumes that all HSCs are equally likely to divide or differentiate. Although the model explains the loss of clonal diversity between six months and four years post transplantation, it does not account for the delayed emergence of previously undetected clones even after four years. Moreover, the one-compartment model is insufficient to fully explain the distribution of clone sizes and the respective changes over time.

These observations led us to hypothesize that HSCs do not constitute a homogeneous pool of cells but rather a dynamic system in which cells can reversibly transition between an active state, characterized by division or differentiation, and a quiescent, potentially regenerative state. This interpretation follows a history of conceptual models that value this dynamic state transitions as an explicit feature of HSC organization, in which the hematopoietic niche plays a crucial role to support HSC quiescence [35, 41, 10, 25, 18]. One key piece of evidence supporting this view is that a homogeneous pool of continuously proliferating and differentiating cells fails to account for the outcomes of label dilution experiments, in particular the observation of long-term label retention [13]. The concept has also been applied to disease settings, such as leukemia, in which this tight regulation between HSCs and their regulating niches is disturbed, providing mutated clones with a competitive advantage [34, 11, 16, 15].

Here we propose a stochastic, two-compartment model of clonal hematopoiesis where HSCs can transition between an active cycling phase and a quiescent state within the niche. Testing this model against a simpler one-compartment model, we find that hematopoietic reconstitution involves not only stochastic fate decisions but also regulated transitions between quiescence and activation. Introducing a second compartment provides a mechanistic explanation for the persistence and delayed activation of certain clones, aligning better with long-term tracking data from the respective NHP models. The quantitative evidence for the existence of a quiescent compartment also suggests that interactions within the stem cell niche play a crucial role in long-term hematopoiesis.

Methods

Data

In a 2023 study, Radtke et al. [33] investigated clonal dynamics following autologous HSC transplantation after full-body irradiation in pigtail macaques, tracking the process at the clonal level from early recovery to long-term engraftment. Their high-density sampling approach uses genetic barcodes to produce a rich dataset containing temporal clonal information for two monkeys spanning over four years of follow-up.

Using this data available in the GitHub repository “ISA-Clone-tracking” [dataset][19], we extracted the clonal time courses of the two animals (Z13264 and Z14004). We restricted our attention to unsorted white blood cells (WBC) gathered from the peripheral blood, since this is the most complete and consistent data source over the whole experimental period

(Supplementary Figure S1). We assume that the clonal composition of WBCs is a sufficient approximation of clonality at the HSC level.

Clonality Metrics of the HSCs

From the extracted data, we derived three metrics reflecting different aspects of HSCs clonal dynamics (Figure 1):

- C*lonal diversity* is an approximation of the number of HSC clones present in the recipient animal at a given time point. Since there are no further transplantations inducing additional clones during the follow up, the number of clones progressively declines over time. Therefore, clonal diversity at a given time point refers to the number of remaining clones that are measured at this time point or any time point thereafter.

- S*ingle-time occurring (SO) clones* are those exclusively observed at a single time point. At a given time point, the *SO fraction* is defined as the number of SO clones divided by the total number of all observable clones. This metric approximates the abundance of small clones that are likely lost over time.

- The *relative clone size distribution* describes the pattern of clonal abundances observed in the marked population. At each time point, the read count of a clone is divided by the total number of reads to obtain its relative abundance. Clones are then grouped by abundance on a logarithmic scale, and the resulting histogram provides an empirical estimate of the overall clone size distribution.

Samples taken on day 249 for Z13264 and on day 138 for Z14004 are remarkable as they show an increased sequencing depth and detect a higher number of especially small clones. This manifests by the peak of SO clones in Figures 1B and the corresponding sharp drop in clone diversity in Figure 1A.

Stochastic Models

A defining characteristic of hematopoietic stem cells (HSCs) is their dual ability to continuously generate differentiated blood cells while simultaneously sustaining their own population through self-renewal. Based on this principle, a simple stochastic model of HSC dynamics can be constructed, where cells either proliferate or differentiate at defined rates. As highlighted by Radtke and colleagues [33], a model based on these two mechanisms is sufficient to account for the progressive loss of clonal diversity over time. However, in order to account for other long-term phenomena such as label dilution kinetics [41, 2], it has been suggested that HSCs transition between different activity levels, namely an active state, in which they can proliferate and differentiate, and a predominantly quiescent state [35, 13, 25].

Here, we apply and quantitatively compare two different model structures, representing these distinct assumptions. In order to account for clonal identity, we formulate the model in a stochastic framework in which the cells can perform different actions with given propensities.

First, we consider a one-compartment model for the HSC pool, which is subdivided into n clonal subpopulations. Each clone proliferates at a maximal rate p_A and differentiates with rate d_A. Proliferation is density-regulated by a logistic term that depends on the carrying capacity k_A and the total number of HSCs $N^A = \sum_{i=1}^{n} N_i^A$, where N_i^A represents the abundance of clone i. At each time step, a cell can either proliferate, giving rise to two new HSCs of the same clone, or differentiate into a progenitor cell. The resulting proliferation and differentiation rates define the propensities of each event, and the likelihood of each event occurring is proportional to its corresponding propensity, as summarized in Table 1.

We also consider a two-compartment model that extends the one-compartment model by considering an additional compartment of quiescent HSCs, with clonal abundance N_i^Q for each clone. The transition between the active and quiescent states is described by density-regulated transition rates t_{AQ} and t_{QA} for deactivation and activation, respectively. Additionally, a carrying capacity k_Q for the quiescent compartment is considered. In this new model, the transition of active cells into the quiescent state is dependent on the base transition rate and the free-space in the target compartment, while quiescent cells have the ability to activate and return to the active compartment in an analogous way (Table 1). This model structure mirrors the one proposed by Hoffmann and colleagues [16], but is extended to accommodate an arbitrary number of clones with identical propensities.

A schematic representation of the second model is shown in Figure 2. It is straightforward to see that the first model is a special case of the second, obtained by removing activation and deactivation, i.e. setting $t_{AQ} = t_{QA} = 0$. Furthermore, the one compartment model is equivalent to that proposed by Radtke and colleagues [33] when proliferation and differentiation have the same propensity.

	Action	Propensity
Active Cells	Proliferation	$p_A\left(1 - \frac{\Sigma A_i}{k_A}\right)$
	Differentiation	d_A
	Deactivation	$t_{AQ}\left(1 - \frac{\Sigma Q_i}{k_Q}\right)$
Quiescent Cells	Activation	$t_{QA}\left(1 - \frac{\Sigma A_i}{k_A}\right)$

Table 1: Propensities for different cellular actions in the stochastic models. Each action occurs with a probability proportional to its propensity, which is determined by a base rate and the carrying capacity of the target compartment. The one-compartment model considers only active cells and actions of proliferation and differentiation, while the two-compartment model considers both cell states and all actions.

Data for model fitting

The dataset from Radtke et al. [33] provides the clonal profile of two animals over a period of almost four years. We acknowledge that the early post-transplantation phase is characterized by dynamic fluctuations and transient contributions from short-lived, lineage-committed progenitors. In humans, it has been shown that these cells dominate hematopoiesis for several months, while long-term HSCs only gradually take over full hematopoietic functionality, typically after 6 to 12 months [3]. Likewise, studies in pigtailed macaques have shown that the hematopoietic system stabilizes over a period of 3 to 12 months before reaching homeostasis [32]. To avoid confounding short-term effects and capture more stable, long-term HSC-driven hematopoiesis, we focus our model analysis on time points beyond 360 days.

Moreover, as we are working with stochastic models, we aim to limit our analysis to more reliable measurements with lower intrinsic variability. For the SO clone fraction metric, we exclude the sample of animal Z13264 at day 532, as it only contains information on 17 cells. At such a small sample size even minor stochastic fluctuations across simulations tend to heavily impact this metric. Similarly, for the relative clone size distribution, we focus our fitting routine on small and medium-sized clones, with a relative size of around 10^{-4} and 10^{-3}. In the data, these clones typically account for more than 85% of the total number of clones, providing a robust basis for model comparison. Evaluated on a logarithmic scale, these abundant clones are less influenced by stochastic fluctuations, whereas the rarer large clones (relative size 10^{-3}~10^{-2}) display higher variability across simulations with the same parameters.

Furthermore, fitting of the clone size distribution is limited to two representative time points per animal (days 397 and 462 for Z13264, and 362 and 495 for Z14004). These have been selected based on high sequencing depth and temporal proximity to the beginning of the homeostatic phase. The limitation to two time points only is warranted to prevent an over-representation of this metric compared to the clonal diversity and the SO clones.

In Supplementary Table S1 we present a summary of the available data points and those used for model fitting.

Simulation Routines and Optimization

In order to derive parameter configurations for each model class that optimally describe the experimental data, we minimize the squared sum of the distance between model and data on a logarithmic scale (residual sum of squares, RSS) across all metrics. To generate model predictions, we employ a customized R-routine that simulates clonal dynamics and the process of blood sampling at specified time points. Technically, the simulation proceeds as follows: given the number of clones and the rates described in the previous section, at each time-step we determine how many cells from each clone will perform a possible action, based on their propensities (Table 2). The carrying capacity is fixed at 10^5 cells per compartment in the two-compartment setting, and at $2*10^5$ (active) cells in the single-compartment case (see Cosgrove et al. [7] and Mitchel et al. [24] for reference values in mice and men). The corresponding compartments are then updated accordingly. At time points corresponding to when blood was sampled in the *in-vivo* experiment, an equal

number of cells is randomly chosen from the active compartment for clonal analysis. Tracking these sampled pools over time allowed us to derive temporal representations of clonal diversity, relative clone size distributions and the fraction of single-time occurring clones.

For the optimization, we compare the model's predictions with the actual data and quantify the difference in terms of the residuals for each data point (formally, the squared distance between model and data on logarithmic scale). Since the comparison is based on all three available metrics, we first normalize the residuals by the observed variation at each particular data point (obtained by repeatedly simulating with a previously defined parameter set), and then average the normalized residuals within each metric. Finally, the difference is calculated as the sum of the those residuals across all three metrics. This approach balances the contribution of each metric, ensuring that no single one disproportionally influences the fitting process.

For a given set of model rates, this approach returns a numerical value indicating the quality of the model's fit to the data. We use this measure within a genetic algorithm [17] to explore the parameter space and identify optimal parameter sets for the initial number of clones and the event rates (p_A, d_A, t_{AQ}, t_{QA} and p_Q, see Supplementary Table S2). Technically, we use the GA package in R [37], which at each generation (iteration), generates a population of instances (i.e. set of parameters) which are evaluated with respect to the above quality measure. The best-performing instances (parameter sets within the population) are then used to generate the next generation, thereby gradually improving the fit over successive iterations. The process continues until convergence is reached, defined as the point where no further improvement in the fit is observed over several generations.

Model Evaluation Metrics

For the optimization routine we use the normalized and averaged distance measure outlined above, which can be viewed as a generalized Residual Sum of Squares (RSS) on the log scale. For models with the same number of parameters, lower RSS values indicate better fit. To *compare different models* with different numbers of parameters, we additionally use the Akaike Information Criterion (AIC) [1] and the Bayesian Information Criterion (BIC) [36]. Both criteria incorporate penalties for model complexity based on the number of parameters, thereby helping to identify models that balance goodness-of-fit with parsimony. Lower AIC and BIC values indicate a better fit.

Results

Clonal convergence in a primate model

Tracking hematopoietic clonal contributions across species, including mice [12, 23, 6], non-human primates [32, 20], and humans [3], has revealed rich dynamic behaviors. Following clonal marking events, some clones expand while others contract, resulting in a gradual loss of clonal diversity over time. This phenomenon occurs even in the absence of selective advantages and reflects the inherent stochastic competition among individual cells.

The term clonal conversion refers to this dynamic of reshaping the clonal landscape and the sequential loss of clones over time.

In a recently published dataset [33], Radtke and colleagues investigated the long-term behavior of genetically marked hematopoietic clones in pigtailed macaques (Figure 1). This comprehensive dataset allows deriving different diversity metrics, enabling detailed quantification of clone size distributions over extended time periods. It is interesting to see that clonal diversity drops rapidly in the early post-transplant phase, most likely as a result of the transient contributions from short-lived, lineage-committed progenitors (Figure 1A). However, on longer time scales, clonal diversity only decreases slowly. Similarly remarkable is the observation that even after several years post transplantation there is an occurrence of previously undetected clones that only occur once (Figure 1B). Changes in clone size distribution indicate that larger clones become more dominant while there is still a substantial fraction of smaller clones remaining (Figure 1C, D).

Model based description of the data using a one-compartment model

To assess how well different mechanistic assumptions capture the outlined clonal dynamics, we systematically evaluated the performance of a one- versus a two-compartment model.

The one-compartment model represents a simplified view of hematopoiesis, where all HSCs remain in an homogeneous state, in which they continuously proliferate and differentiate. This model is equivalent to the one presented by Radtke et al. [33] and provides a reasonable fit for the overall measure of clonal diversity. However, when attempting to fit all metrics combined, the overall fit deteriorates significantly. Notably, in order to reproduce even a minimal fraction of SO clones beyond the one-year horizon (Figure 3B), the model requires an unrealistically large number of initial clones. This, in turn, leads to an overestimation of total clone numbers (Figure 3A) and a failure to reproduce the observed clone size distributions (Figure 3C–D). Even under these conditions, the predicted SO clone fraction remains well below experimental values, which consistently show the appearance of new clones even at later time points.

Optimizing this model with a lower emphasis on the SO fraction yields visually improved fits, but still shows substantial discrepancies. Even the best fit achieved with this modification fails to capture the abundance of small clones, which appear on the left tail of the clone size distributions (Supplementary Figure S2). This limitation also affects the model-based approximation of the clonal diversity, leading to an underestimation of the overall number of clones.

Conceptually, these results highlight a key limitation of the one-compartment model: considering only one homogeneous HSC population, causes small clones to be more vulnerable to the competitive pressure from larger clones, making them unable to persist over long periods. This limitation suggests that additional dynamics are necessary to fully explain the observed clonal behavior.

Model based description of the data using a two-compartment model

We propose that an intrinsic HSC heterogeneity, modeled as a transient quiescent state, improves the model predictions by explaining long-term clone persistence and the sporadic occurrence of small clones. To test this hypothesis, we introduce an additional compartment and the possibility to transition between active and quiescent states.

In the two-compartment model, HSCs are able to reversibly transition between an active state, where they proliferate and differentiate, and a quiescent state, where they remain proliferatively inactive. Analyzing the clonal profile generated by this model, we confirm that the inclusion of the quiescent state yields the expected results. Using this model, we were able to capture the continued emergence of SO clones throughout the entire sampling period (Figure 3F). Moreover, both the overall number of clones (Figure 3E) and their contributions to the HSC pool (Figure 3G and 3H) were in close agreement with the experimental data.

To quantify and compare the goodness of fit for each model, we evaluated the residual sum of squares (RSS), the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). As shown in Supplementary Figure S3, the RSS for the two-compartment model is consistently lower compared to the one-compartment model across all evaluated metrics, confirming the visual intuition that the two-compartment model provides a significantly better fit to the data. Furthermore, both AIC and BIC scores are lower for the two-compartment model, indicating that even after penalizing the additional parameters, the two-compartment model remains better suited to describe the combined metrics (Figure 5 and Supplementary Figure S7).

We still notice an underestimation in the SO clone fraction during an intermediate period for animal Z14004, suggesting that additional mechanisms may be required to fully account for the increased presence of these small clones in this window. Despite this limitation, the two-compartment model successfully reproduces key features of the clonal data, reinforcing the idea that transitions between active and quiescent states are essential for shaping long-term persistence of clonal diversity, appearance and contribution patterns.

Average Time to Events

The frequency of cellular events is a central aspect of biological dynamics, as it defines how often key processes such as division or differentiation occur. Fitting mechanistic models to empirical data enables the estimation of these frequencies or, equivalently, the inverse of it, which is the average time until an event occurs. In computational simulations, these quantities can be measured empirically by monitoring the time required for individual cells to undergo a specific event. As illustrated in Figure 4, the estimated average time for an event depends on the simulation time and stabilizes as it increases well beyond the critical time scale. This convergence behavior, typical of average-based estimations, highlights the importance of longer experiments to accurately capture the true temporal dynamics, especially for infrequent events.

Apart from this computational approach, one can also estimate these average times analytically. The stochastic implementation of the two-compartment model, can be seen as a generalization of the preceding ODE model [16] for the case of several clones with identical

capabilities. In this setting, each cellular action (e.g., proliferation, differentiation, or state transition) occurs with a defined probability per unit time, proportionally to the rate used in the ODE formulation. At homeostasis, the expected time between successive events can be calculated as the inverse of its associated rate, with results summarized in Table 2. Figure 4 shows that the empirically derived average time between events converges toward its analytically predicted counterpart as the simulation time increases.

Evaluating the average time between events for the optimized two-compartment model, we observe that division-related processes (proliferation and differentiation) occur on a much faster timescale compared to state transitions (activation and deactivation). Namely, it takes on average one week for a cell to divide, while it takes on average several years for a cell to activate (Table 2). At first glance, this disparity might suggest that transitions can be neglected. However, omitting them effectively reduces the model to the one-compartment version, which cannot explain long-term persistence of small clones. It is interesting to note that transition times follow an exponential distribution (Supplementary Figure S4), implying that early transition events are still frequent despite long averages. While 50% of the cells transition within the first three and a half years, the fastest 25% of transitions occur within the first one and a half years (Supplementary Table S3). This characteristic of exponential distributions explains why such transitions, though rare in homeostatic situations, substantially influence the overall system behavior.

In summary, although state transitions occur much less frequently than division events, their presence is crucial for the long-term dynamics of the hematopoietic system.

	Event	Average Time (On Equilibrium)	Average Times for the Fitted Two-Compartment Model
Active Cells	Proliferation	$\frac{1}{d_A}$	8 days
	Differentiation	$\frac{1}{d_A}$	8 days
	Deactivation	$\frac{1}{t_{AQ}} + \frac{p_A}{d_A t_{QA}} \frac{k_A}{k_Q}\left(1 - \frac{d_A}{p_A}\right)$	1547 days
Quiescent Cells	Activation	$\frac{p_A}{d_A t_{QA}}$	1863 days

Table 2: Analytical average time to cellular events at homeostatic equilibrium of the Hoffmann model [16]. Each value represents the expected time between two occurrences of a given action for a single cell, assuming the system has reached steady state. Expressions are derived as the inverse of the corresponding action rates in the deterministic ODE model described by Hoffmann et al. Average times for the best-fit parameters for the two-compartment model are shown in the third column.

Different Mechanisms for HSC Proliferation

In the scope of the two-compartment model, the quiescent state is associated with niche regulation that maintains cells in a non-proliferative condition, while active cells divide

independently of any previous actions (niche-independent proliferation). Recent work [26, 15], including our own, has suggested an alternative mechanism in which proliferation is directly linked to the transition from quiescence to activation (niche-dependent proliferation). To combine these views, a unified model has been proposed that incorporates both processes: in addition to random divisions of active cells, a fraction p_Q of cells exiting quiescence immediately undergo proliferation [15]. Notably, the two-compartment model discussed here can be perceived as a particular case of this unified model with $p_Q = 0$ (see Supplementary Figure S5). We raise the question whether this unified model is better suited to explain the long latency of hematopoietic clones after transplantation.

Incorporating the additional parameter p_Q in our fitting routine for the unified model, we observe that it achieves a fit similar to that of the earlier two-compartment model. Both reproduce key trends across all metrics with comparable accuracy (Supplementary Figure S6). This similarity is also reflected in overlapping ranges of the RSS (Figure 5A). Model selection criteria such as AIC and BIC (Figure 5B and Supplementary Figure S7) further reinforce this interpretation, showing only small differences between the two models. This suggests that niche-dependent proliferation may not be robustly identified from the shape of clonal dynamics under steady-state conditions, while a more prominent role in perturbed or stressed hematopoietic environments should not be excluded. However, comparing these metrics with the corresponding one-compartment model further supports the hypothesis that niche influence, whether through a quiescent compartment or via its role in proliferation, is pivotal to explain the observed dynamics.

Discussion

In this study, we employed high-resolution clonal tracking of hematopoietic cells in macaques to demonstrate that an additional layer of heterogeneity among HSCs is necessary to particularly explain the late emergence of smaller clones. Precisely, referring to a previously established concept, in which HSCs reversibly transit between a proliferatively inactive/quiescent state and an active, proliferative state, clearly outperforms simpler models that consider HSCs as a single, homogeneous population. Using a set of particular metrics to describe key features of the temporal clonal data set, we show that while a model incorporating only proliferation and differentiation can reproduce the gradual loss of clonal diversity, it fails to capture the observed distribution of clone sizes and entirely misses the long-term persistence of small clones. Introducing heterogeneity into the HSC population by distinguishing between actively cycling cells and niche-bound quiescent cells, yielded a remarkably improved fit to all metrics. Our results confirm that the addition of a quiescent compartment allows for a sustained coexistence of small and large clones and, in particular, the long-term emergence of single-time occurring clones across the observation period, which is in good agreement with experimental observations.

These findings highlight the importance of acknowledging a duality in the appearance of HSCs, to be either in an active or quiescent state. This concept has been long discussed on a theoretical level [22, 13], and is increasingly gaining recognition to explain biological data [41, 10, 40, 25, 18]. A prominent example arises from label dilution experiments [41, 25] where the biphasic decline of label-retaining HSCs, characterized by an initial rapid loss followed by a slower secondary phase, can only be explained by assuming that a substantial subset of HSCs exhibits markedly slower turnover dynamics.

It is an open conceptual question whether quiescence and activation in HSCs are truly two distinct and discrete states, or whether they are better understood as two extremes along a continuous spectrum of functional activity. In the latter view, individual HSCs may gradually shift their proliferative activity, potentially in response to external cues, but also to spatial and metabolic conditions. This would be in contrast to a sharp transition between binary states, as we consider it in our analysis. However, the two-state model offers conceptual clarity, is easier to represent in terms of mathematical or conceptual models, and also aligns well with experimentally accessible features such as flow cytometry based cell classification. In turn, it may only represent a simplified abstraction of a more complex biological reality.

While the division rates estimated for the two-compartment model are compatible with other data sources [41, 13, 5, 4, 3, 7], we notice that this model describes the observed data with fewer clones compared to the one-compartment model. It appears that the latter model structure tries to compensate for the rapid loss of clones by a higher initial clone number. Furthermore, while transitions between active and quiescent states emerge as a crucial mechanism to explain the data, their rate is much lower than the cell division rate. Still, the model fits clearly indicate that the slow but continuous flux of cells between compartments is a requirement for the steady renewal of the active pool, thereby accounting for the persistent emergence of new clones even at late sampling times.

Although our model provides a good overall fit to the data, some aspects remain unexplained. For instance, it cannot simultaneously reproduce the rapid loss of diversity observed shortly after transplantation and the slower dynamics that dominate during later phases. Moreover, while the model predicts the existence of single-occurring clones, it fails to reproduce their observed frequency at intermediate time points. These limitations suggest the need for more refined models that incorporate, for instance, heterogeneity in clonal capacities or context-dependent regulatory mechanisms. Importantly, such extensions can be naturally accommodated within the same two-compartment framework, offering a clear path forward for future work.

In summary, our analysis demonstrates that describing HSCs as a heterogeneous population comprising two complementary functional activity states provides a suitable explanation for the resulting clonal dynamics. It supports the notion of reversible HSC quiescence as a key element to understand long-term maintenance of clonal diversity. Our approach highlights the potential of using clonality data to explore the underlying mechanisms of HSC interaction and regulation, enabling the simultaneous evaluation of several different metrics in a unified framework and leading to more robust models.

Author Contribution

YGV: Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.
LT: Methodology, Data Curation, Writing - Review & Editing.
ACF, IG: Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.

Acknowledgments

YGV was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. ACF was supported by Alexander von Humboldt Foundation and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and partially supported by FAPEMIG RED-00133-21.

Competing interests

All authors declare no financial or non-financial competing interests.

Data and Code Availability

Data and code available on https://github.com/YGVilela/HSC.

References

1. Akaike, H. (1998). *Information Theory and an Extension of the Maximum Likelihood Principle* (pp. 199–213). https://doi.org/10.1007/978-1-4612-1694-0_15
2. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H., & Moore, K. (2016). Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell*, *167*(5), 1296-1309.e10. https://doi.org/10.1016/j.cell.2016.10.022
3. Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M. P., Giannelli, S., Neduva, V., Dow, D. J., Schmidt, M., von Kalle, C., Roncarolo, M. G., Ciceri, F., Vicard, P., Wit, E., … Aiuti, A. (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell*, *19*(1), 107–119. https://doi.org/10.1016/j.stem.2016.04.016
4. Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., & Rodewald, H.-R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, *518*(7540), 542–546. https://doi.org/10.1038/nature14242
5. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P., & Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells in vivo. *Blood*, *117*(17), 4460–4466. https://doi.org/10.1182/blood-2010-08-303537
6. Cornils, K., Thielecke, L., Hüser, S., Forgber, M., Thomaschewski, M., Kleist, N., Hussein, K., Riecken, K., Volz, T., Gerdes, S., Glauche, I., Dahl, A., Dandri, M., Roeder, I., & Fehse, B. (2014). Multiplexing clonality: combining RGB marking and genetic barcoding. *Nucleic Acids Research*, *42*(7), e56. https://doi.org/10.1093/nar/gku081
7. Cosgrove, J., Hustin, L. S. P., de Boer, R. J., & Perié, L. (2021). Hematopoiesis in numbers. *Trends in Immunology*, *42*(12), 1100–1112. https://doi.org/10.1016/j.it.2021.10.006
8. Cullen, S. M., Mayle, A., Rossi, L., & Goodell, M. A. (2014). Hematopoietic stem cell development: an epigenetic journey. *Current Topics in Developmental Biology*, *107*, 39–75. https://doi.org/10.1016/B978-0-12-416022-4.00002-0

9. Ema, H., Morita, Y., & Suda, T. (2014). Heterogeneity and hierarchy of hematopoietic stem cells. *Experimental Hematology*, *42*(2), 74-82.e2. https://doi.org/10.1016/j.exphem.2013.11.004
10. Essers, M. A. G., Offner, S., Blanco-Bose, W. E., Waibler, Z., Kalinke, U., Duchosal, M. A., & Trumpp, A. (2009). IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*, *458*(7240), 904–908. https://doi.org/10.1038/nature07815
11. Fassoni, A. C., Baldow, C., Roeder, I., & Glauche, I. (2018). Reduced tyrosine kinase inhibitor dose is predicted to be as effective as standard dose in chronic myeloid leukemia: a simulation study based on phase III trial data. *Haematologica*, *103*(11), 1825–1834. https://doi.org/10.3324/haematol.2018.194522
12. Gerrits, A., Dykstra, B., Kalmykowa, O. J., Klauke, K., Verovskaya, E., Broekhuis, M. J. C., de Haan, G., & Bystrykh, L. v. (2010). Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood*, *115*(13), 2610–2618. https://doi.org/10.1182/blood-2009-06-229757
13. Glauche, I., Moore, K., Thielecke, L., Horn, K., Loeffler, M., & Roeder, I. (2009). Stem cell proliferation and quiescence--two sides of the same coin. *PLoS Computational Biology*, *5*(7), e1000447. https://doi.org/10.1371/journal.pcbi.1000447
14. Greenbaum, A., Hsu, Y.-M. S., Day, R. B., Schuettpelz, L. G., Christopher, M. J., Borgerding, J. N., Nagasawa, T., & Link, D. C. (2013). CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature*, *495*(7440), 227–230. https://doi.org/10.1038/nature11926
15. Harnisch, C., Glauche, I., Uschner, F., Roeder, I., & Fassoni, A. C. (2024). *Mechanisms of leukemic stem cell interaction in AML: insights from complementary mathematical modeling approaches.* [preprint] https://doi.org/10.21203/rs.3.rs-3958767/v1
16. Hoffmann, H., Thiede, C., Glauche, I., Bornhaeuser, M., & Roeder, I. (2020). Differential response to cytotoxic therapy explains treatment dynamics of acute myeloid leukaemia patients: insights from a mathematical modelling approach. *Journal of the Royal Society, Interface*, *17*(170), 20200091. https://doi.org/10.1098/rsif.2020.0091
17. Holland, J. H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.
18. Ibneeva, L., & Grinenko, T. (2024). Dissecting dormancy and quiescence in hematopoietic stem cells. *Frontiers in Hematology*, *3*. https://doi.org/10.3389/frhem.2024.1401713
19. KiemLab-RIS. (2022). ISA-Clone-tracking [dataset]. GitHub. https://github.com/KiemLab-RIS/ISA-Clone-tracking
20. Kim, S., Kim, N., Presson, A. P., Metzger, M. E., Bonifacino, A. C., Sehl, M., Chow, S. A., Crooks, G. M., Dunbar, C. E., An, D. S., Donahue, R. E., & Chen, I. S. Y. (2014). Dynamics of HSPC repopulation in nonhuman primates revealed by a decade-long clonal-tracking study. *Cell Stem Cell*, *14*(4), 473–485. https://doi.org/10.1016/j.stem.2013.12.012
21. 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., & Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, *561*(7724), 473–478. https://doi.org/10.1038/s41586-018-0497-0

22. Loeffler, M., & Roeder, I. (2002). Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models--a conceptual approach. *Cells, Tissues, Organs*, *171*(1), 8–26. https://doi.org/10.1159/000057688
23. Lu, R., Neff, N. F., Quake, S. R., & Weissman, I. L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nature Biotechnology*, *29*(10), 928–933. https://doi.org/10.1038/nbt.1977
24. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K. J., Mende, N., Calderbank, E. F., Jung, H., Mitchell, T., Coorens, T. H. H., Spencer, D. H., Machado, H., Lee-Six, H., Davies, M., Hayler, D., Fabre, M. A., Mahbubani, K., Abascal, F., Cagan, A., Vassiliou, G. S., … Campbell, P. J. (2022). Clonal dynamics of haematopoiesis across the human lifespan. *Nature*, *606*(7913), 343–350. https://doi.org/10.1038/s41586-022-04786-y
25. Morcos, M. N. F., Li, C., Munz, C. M., Greco, A., Dressel, N., Reinhardt, S., Sameith, K., Dahl, A., Becker, N. B., Roers, A., Höfer, T., & Gerbaulet, A. (2022). Fate mapping of hematopoietic stem cells reveals two pathways of native thrombopoiesis. *Nature Communications*, *13*(1), 4504. https://doi.org/10.1038/s41467-022-31914-z
26. Pedersen, R. K., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., Ottesen, J. T., & Stiehl, T. (2023). HSC Niche Dynamics in Regeneration, Pre-malignancy, and Cancer: Insights From Mathematical Modeling. *Stem Cells (Dayton, Ohio)*, *41*(3), 260–270. https://doi.org/10.1093/stmcls/sxac079
27. 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*, 110620. https://doi.org/10.1016/j.jtbi.2021.110620
28. Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., Chen, W., Sauer, S., Wolf, S., Höfer, T., & Rodewald, H.-R. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, *548*(7668), 456–460. https://doi.org/10.1038/nature23653
29. Pereira, A. L., Galli, S., & Nombela-Arrieta, C. (2024). Bone marrow niches for hematopoietic stem cells. *HemaSphere*, *8*(8). https://doi.org/10.1002/hem3.133
30. Perié, L., Hodgkin, P. D., Naik, S. H., Schumacher, T. N., de Boer, R. J., & Duffy, K. R. (2014). Determining lineage pathways from cellular barcoding experiments. *Cell Reports*, *6*(4), 617–624. https://doi.org/10.1016/j.celrep.2014.01.016
31. Pinho, S., & Frenette, P. S. (2019). Haematopoietic stem cell activity and interactions with the niche. *Nature Reviews. Molecular Cell Biology*, *20*(5), 303–320. https://doi.org/10.1038/s41580-019-0103-9
32. Radtke, S., Adair, J. E., Giese, M. A., Chan, Y.-Y., Norgaard, Z. K., Enstrom, M., Haworth, K. G., Schefter, L. E., & Kiem, H.-P. (2017). A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. *Science Translational Medicine*, *9*(414). https://doi.org/10.1126/scitranslmed.aan1145
33. Radtke, S., Enstrom, M., Pande, D., Duke, E. R., Cardozo-Ojeda, E. F., Madhu, R., Owen, S., Kanestrom, G., Cui, M., Perez, A. M., Schiffer, J. T., & Kiem, H.-P. (2023). Stochastic fate decisions of HSCs after transplantation: early contribution, symmetric expansion, and pool formation. *Blood*, *142*(1), 33–43. https://doi.org/10.1182/blood.2022018564
34. Roeder, I., Kamminga, L. M., Braesel, K., Dontje, B., de Haan, G., & Loeffler, M. (2005). Competitive clonal hematopoiesis in mouse chimeras explained by a

stochastic model of stem cell organization. *Blood*, *105*(2), 609–616. https://doi.org/10.1182/blood-2004-01-0282
35. Roeder, I., & 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. https://doi.org/10.1016/s0301-472x(02)00832-9
36. Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2). https://doi.org/10.1214/aos/1176344136
37. Scrucca, L. (2013). **GA** : A Package for Genetic Algorithms in *R*. *Journal of Statistical Software*, *53*(4). https://doi.org/10.18637/jss.v053.i04
38. Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., & Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature*, *514*(7522), 322–327. https://doi.org/10.1038/nature13824
39. Verovskaya, E., Broekhuis, M. J. C., Zwart, E., Ritsema, M., van Os, R., de Haan, G., & Bystrykh, L. v. (2013). Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. *Blood*, *122*(4), 523–532. https://doi.org/10.1182/blood-2013-01-481135
40. Viale, A., & Pelicci, P. G. (2009). Awaking stem cells from dormancy: growing old and fighting cancer. *EMBO Molecular Medicine*, *1*(2), 88–91. https://doi.org/10.1002/emmm.200900019
41. Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., Macdonald, H. R., & Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, *135*(6), 1118–1129. https://doi.org/10.1016/j.cell.2008.10.048

Figures

Figure 1

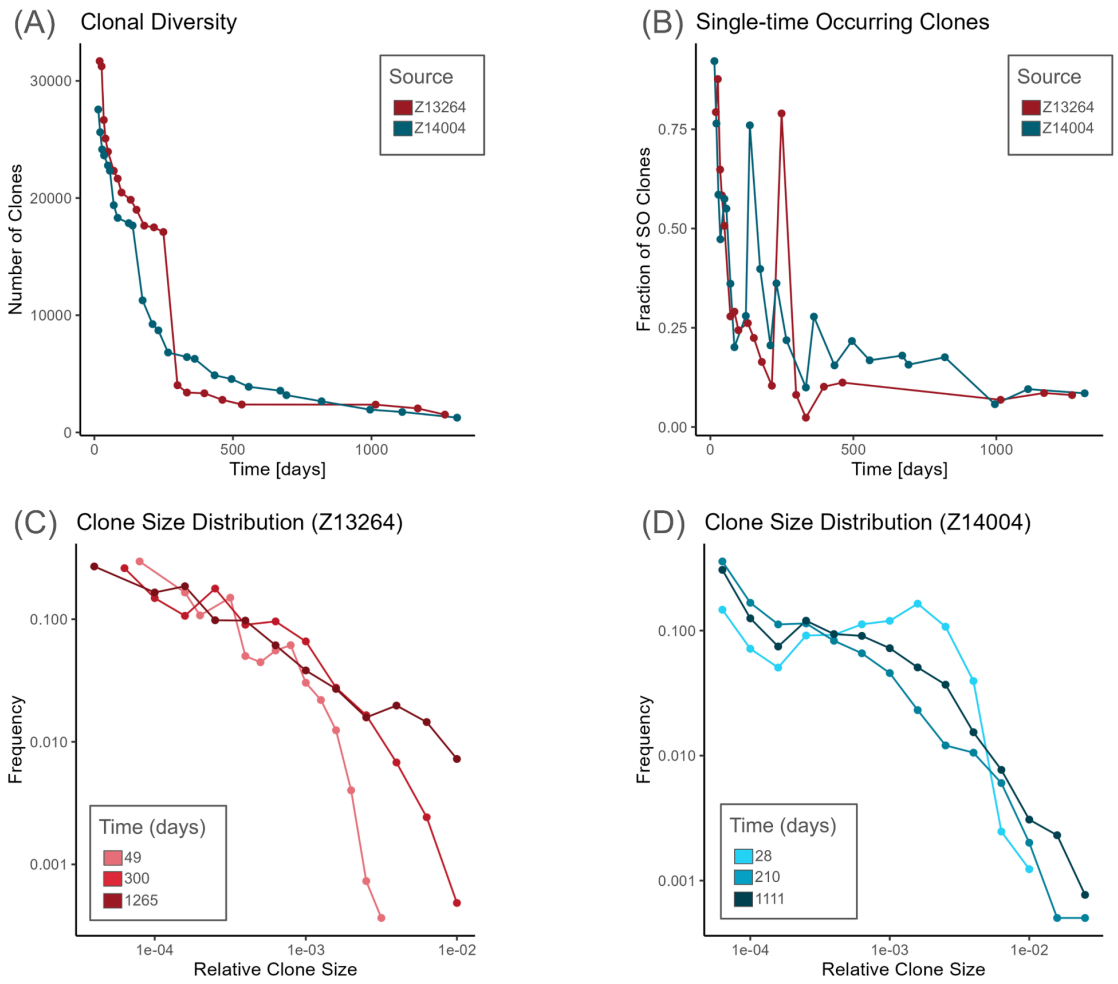

Figure 1 - Metrics of HSC clonality: Metrics of HSC clonality, as derived from Radtke et al. [33] (A) Clone diversity, measured as the number of distinct clones detected over time. (B) Fraction of single-time occurring (SO) clones shown at each time point, i.e., the number of clones observed only at that time point divided by the total number of clones observed on that time point. (C-D) Clone size distribution for animals Z13264 and Z14004, respectively, shown as histograms of clone relative abundance (clone size divided by total of reads) on a log-log scale. Dots indicate observed data; lines connect observations for clarity. Colors denote animal id and sampling time as indicated in the legend.

Figure 2

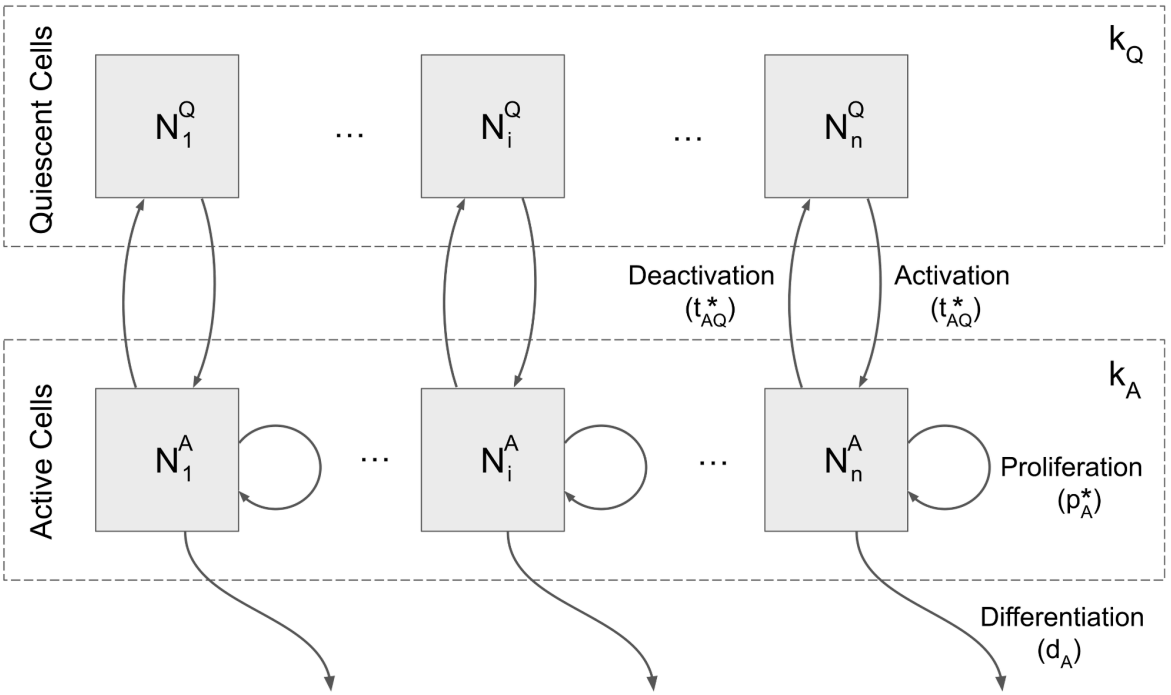

Figure 2 - Two-compartment model of HSCs dynamics: Dashed boxes indicate the "active" and "quiescent" compartments with their respective carrying capacities, k_A and k_Q, and one subpopulation for each clone represented as gray boxes. N_i^A and N_i^Q indicate the abundance of clone i in the active and quiescent compartments, respectively. Arrows indicate the flux of cells when the corresponding action is taken with propensities as defined in Table 1. A one-compartment model can be obtained by removing the quiescent compartment along with the actions that interact with it.

Figure 3

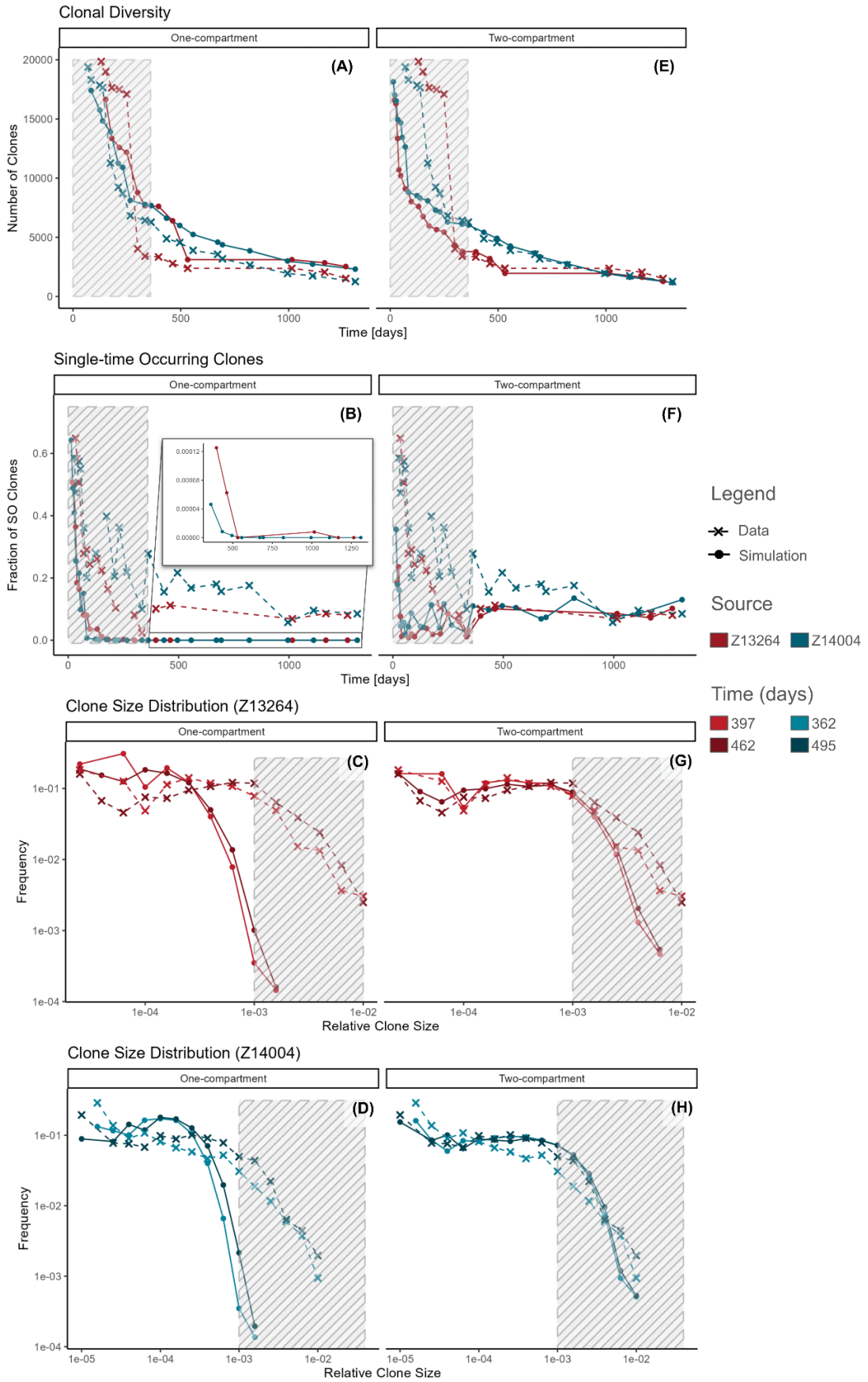
Clonal Diversity
One-compartment
Two-compartment
(A)
(E)
Number of Clones
Time [days]
Single-time Occurring Clones
(B)
(F)
Fraction of SO Clones
Clone Size Distribution (Z13264)
(C)
(G)
Frequency
Relative Clone Size
Clone Size Distribution (Z14004)
(D)
(H)
Legend
Data
Simulation
Source
Z13264
Z14004
Time (days)
397
362
462
495

Figure 3 - Model fits for the one vs two-compartment models: Best fit for the one (A-D) and two-compartment models (E-H) optimized with respect to the metrics derived from data by Radtke et al. [33]. Dots indicate model predictions with lines connecting observations for clarity; experimental data is indicated by **x** marks. Striped areas indicate data which is not used for fitting. Model parameters available in Supplementary Table 4.

Figure 4

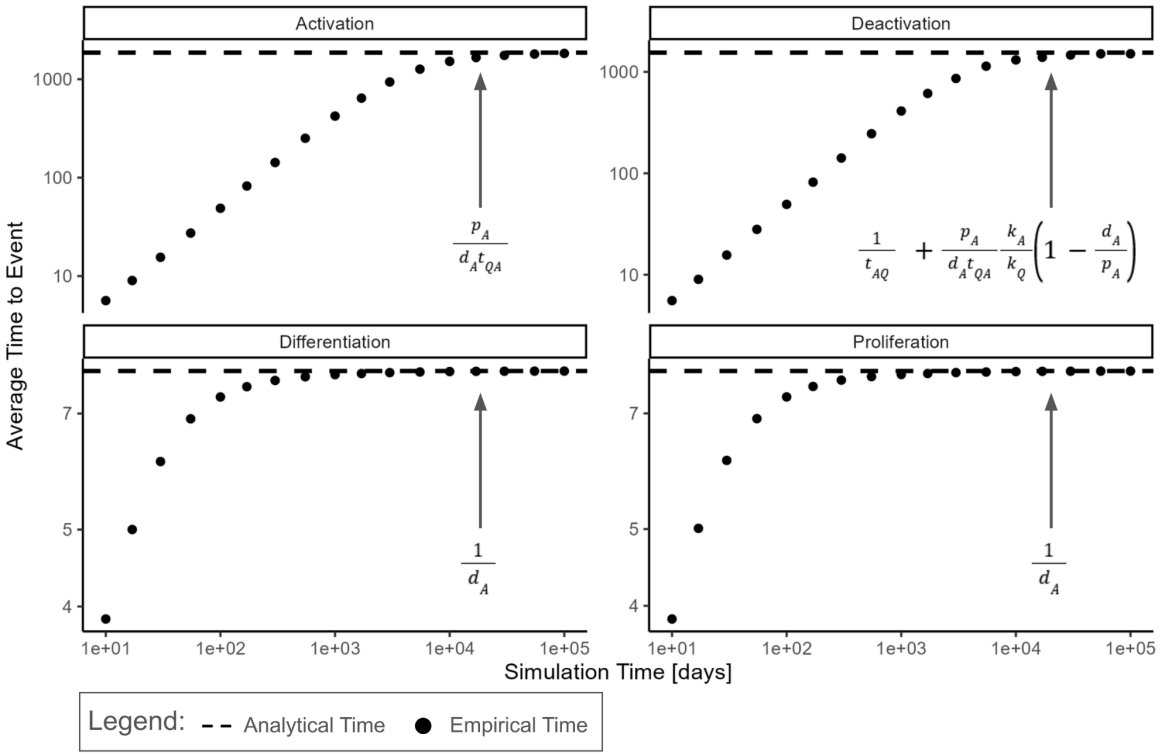

Figure 4 - Average time to events: Empirical and analytical estimates for the average time to event at homeostatic equilibrium. Average times are estimated for the optimal parameter configuration obtained for the two-compartment model (Supplementary Table 4). Dots indicate the estimated average time between events recorded in the simulations at different time points. As the simulation period increases, the empirical estimates (dots) converge toward its analytically predicted counterpart (dashed lines).

Figure 5

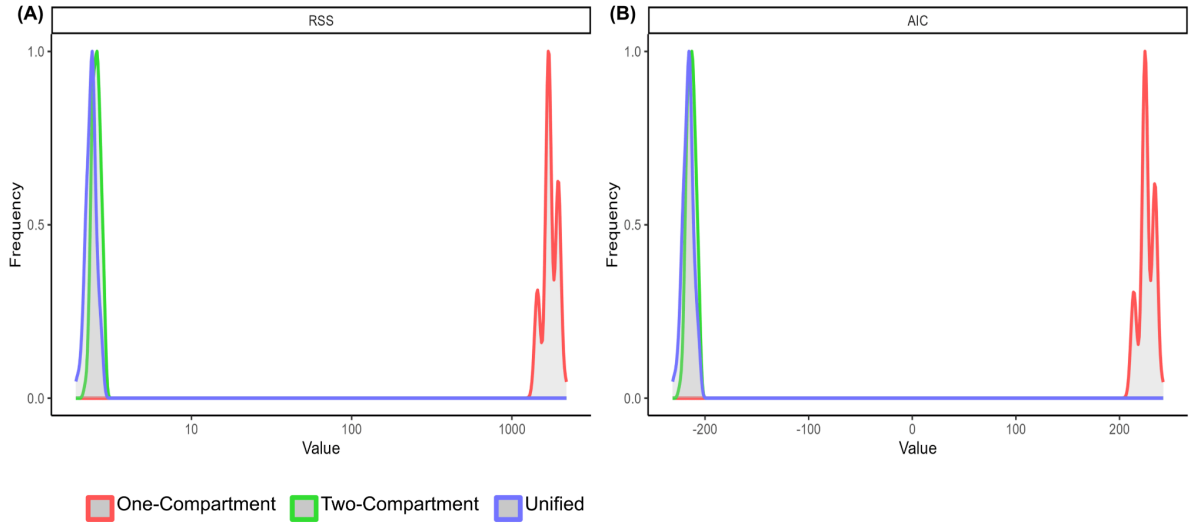

Figure 5 - Model evaluation metrics: Distribution of Residual Sum of Squares (RSS) and Akaike Information Criterion (AIC) values from 100 simulations using the best-fit parameters (see Supplementary Table 4) of the one-compartment (red), two-compartment (green) and unified models (blue).