

RESEARCH ARTICLE

A stochastic model of myeloid cell lineages in hematopoiesis and pathway mutations in acute myeloid leukemia

Frank Jäkel^{1*}, Oliver Worm¹, Sascha Lange¹, Roland Mertelsmann²

¹ Psiori GmbH, Freiburg, Germany, ² Innere Medizin I, Universitätsklinikum Freiburg, Freiburg, Germany

* frank@psiori.com



OPEN ACCESS

Citation: Jäkel F, Worm O, Lange S, Mertelsmann R (2018) A stochastic model of myeloid cell lineages in hematopoiesis and pathway mutations in acute myeloid leukemia. PLoS ONE 13(10): e0204393. <https://doi.org/10.1371/journal.pone.0204393>

Editor: Ying-Mei Feng, Beijing Key Laboratory of Diabetes Prevention and Research, CHINA

Received: July 17, 2017

Accepted: September 7, 2018

Published: October 1, 2018

Copyright: © 2018 Jäkel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Work is purely theoretical. Where data is discussed the relevant information is in the paper.

Funding: Funding was partly provided by the private Addison Fischer Institute for Research and Ethical Use of Artificial Intelligence through a grant to PSIORI GmbH, Freiburg. FJ and OW were employees of PSIORI at the time the research was conducted. SL is the CEO of PSIORI. There was no additional external funding received for this study. The funders had no additional role in study design,

Abstract

A model for hematopoiesis is presented that explicitly includes the erythrocyte, granulocyte, and thrombocyte lineages and their common precursors. A small number of stem cells proliferate and differentiate through different compartments to produce the vast number of blood cells needed every day. Growth factors regulate the proliferation of cells dependent on the current demand. We provide a steady state analysis of the model and rough parameter estimates. Furthermore, we extend the model to include mutations that alter the replicative capacity of cells and introduce differentiation blocks. With these mutations the model develops signs of acute myeloid leukemia.

Introduction

All blood cells are generated from very few stem cells and go through several stages of cell division and differentiation that greatly amplify the number of cells. In fact, one cell division per day at the stem cell stage is thought to lead to roughly 350 billion cells flowing out into the blood stream every day. How is this massive amplification achieved? And how does this process explain the dynamical changes in blood cell counts that clinicians observe in their daily work, e.g. in leukemia?

There is a long history of mathematical modeling of hematopoiesis with two traditions, one rooted in differential equations and one in stochastic modeling [1, 2]. The dynamical and control-theoretic aspects of hematopoiesis are naturally captured with differential equations. In contrast, the detailed biology of cell proliferation and differentiation is often easier to model with discrete stochastic processes, which often go down to the single cell level and sometimes even include genetic and other intracellular processes. This tension between single cell models and models of the global dynamics is in no way unique to hematopoiesis, it exists in all areas of systems biology.

However, a specific challenge in hematopoietic modeling is that the whole system crucially depends on a very small number of hematopoietic stem cells, making it very desirable to have models that span the micro- and the macro-level [3]. Also, biomedical research on the pathologies of the hematopoietic system increasingly focuses on molecular and genetic explanations. For example, the genes that are associated with human myeloid leukemia are extremely well

data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of the authors are articulated in the 'author contributions' section.

Competing interests: FJ, OW, and SL work for a data science company, PSIORI GmbH, Freiburg. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

characterized [4–7] and cancerogenesis, in general, is now understood as arising from a very small number of mutations in a variety of pathways that tightly regulate cell proliferation and cell death [8–10]. These genetic and molecular insights can be incorporated into models of the global dynamics [11, 12] but without modeling single cells the effects of single mutations on leukemogenesis cannot be studied directly.

Here, we present a stochastic, compartmental model that counts single cells at various stages of hematopoiesis. Our model is strongly inspired by the model of Dingli et al. [13] that was later generalized and analyzed in detail by Werner et al. [14]. In the original model no distinction between different cell types is made and hence the different characteristics of, for example, the erythrocyte, granulocyte, and thrombocyte lineages in hematopoiesis cannot be taken into account. The major extension we propose here is to explicitly model these three myeloid lineages of hematopoiesis. In addition, we will also include a feedback mechanism with lineage-specific growth factors. As we account for the three lineages and their common precursors the feedback mechanisms that we propose is much more detailed than previous extensions of the original model that also included feedback [15]. Furthermore, setting the parameters of our model to realistic values is harder than in the original model because of interactions between the three lineages. We show, however, that rough parameter estimates can still be obtained by considering the steady state, similar to how Dingli et al. [13] did it. Finally, we extend the model to include single mutations that might account for some aspects of acute myeloid leukemia (AML). In this regard, our model mirrors similar efforts by Werner and colleagues [14, 16, 17], who do not, however, deal with the complications of differentiating between cell lineages.

Methods

Even though our model is based on the model of Dingli et al. [13], the introduction of different cell lineages and the inclusion of cell-lineage specific growth factors make it easier to explain our model from scratch, rather than to present it as an extension of the original model. This is what we will do in the *Methods* section. The *Results* section will then give a theoretical analysis of the new model and show that based on this analysis the model's parameters can be set to physiologically plausible values. Finally, we will extend the model slightly to allow for single mutations in single cells and use this extension to simulate the development of acute myeloid leukemia.

A compartmental model

We will consider the numbers of three myeloid types of blood cells: erythrocytes (E), granulocytes (G), and thrombocytes (T). To keep the model simple, we ignore the monocytes and only consider the granulocytes among the leukocytes. Risk of bacterial infections among transplant patients is predominantly correlated with the number of neutrophil leukocytes and therefore the granulocyte count is clinically the more relevant variable. Also, the monocytes only make up a very small proportion of the leukocytes anyway. In addition we will consider the common precursors (C) of the myeloid cells, including the hematopoietic stem cells.

We assume there are K_E , K_G and K_T compartments for the erythrocyte, granulocyte, and thrombocyte lineages. For the common precursors we assume there are $K_C + 1$ compartments with the zeroth compartment being the stem cell compartment. Each compartment has $N_{x,k}(t)$ cells in it at time point t . For example, we assume the zeroth compartment of the common precursors is the stem cell compartment and $N_{C,0}(t)$ is the number of cells in it. $N_{C,K_C}(t)$ is the last stage before cells commit themselves to one of the three lineages. $N_{E,1}(t)$ is then the first stage of the erythrocyte lineage and $N_{E,K_E}(t)$ is the number of erythrocytes in the blood. Likewise for

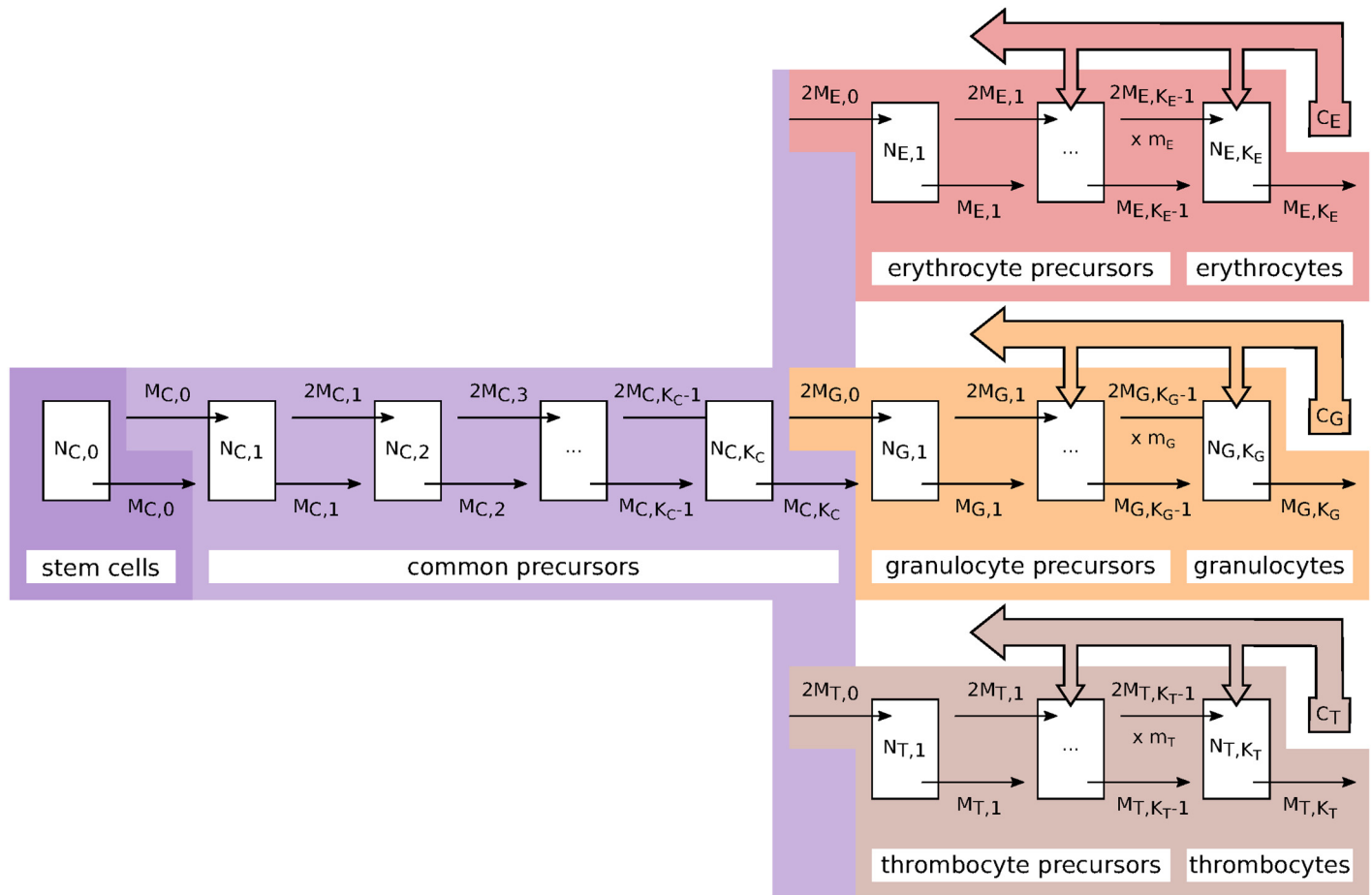


Fig 1. Illustration of the model's compartments. The model consists of $K_C + 1$ compartments for the common precursors. The zeroth compartment is the stem cell compartment. The erythrocyte lineage has K_E compartments, and the granulocyte and thrombocyte lineages have K_G and K_T compartments, respectively. The number of cells in each compartment is given by N and the outflow of each compartment by M with the respective indices. The cells at the final stage (on the very right) model the mature blood cells in the blood stream. Deviations of the respective blood counts from their normal range result in lineage specific growth factors (c_E, c_G, c_T) that feed back to all stages and increase their outflow.

<https://doi.org/10.1371/journal.pone.0204393.g001>

the granulocytes and the thrombocytes. Fig 1 illustrates the compartments of the model. Furthermore, there are a number of growth factors in the bone marrow that change over time. We collect the (log) concentrations for all these growth factors in $c(t) = \{c_E(t), c_G(t), c_T(t)\}$. These growth factors act as a feedback signal to control the outflow of each stage to ensure that the blood counts in the blood stream maintain their normal levels. The rate at which cells leave their compartment therefore depends on these concentrations: $r_{x,k}(c(t))$ measured in 1/days for $x \in \{C, E, G, T\}$.

To simulate the number of cells in each compartment over time we discretize time, $t \in \{j\Delta t \mid j \in \mathbb{N}\}$ measured in days. With Δt sufficiently small the probability that a cell divides itself in any one time step is

$$r'_{x,k}(t) = r_{x,k}(c(t))\Delta t. \tag{1}$$

For $\Delta t \rightarrow 0$ this discrete process becomes an inhomogeneous Poisson process. Since we have $N_{x,k}(t)$ cells in the k th compartment, the number $M_{x,k}(t)$ of the cells that leave this

compartment at time t is a Binomial distribution with

$$M_{x,k}(t) \sim \text{Binomial}(N_{x,k}(t), r'_{x,k}(t)) \tag{2}$$

for $x \in \{C, E, G, T\}$ and all the relevant k .

Stem cell compartment. Cell division in the stem cell compartment is asymmetrical: If a cell divides, it turns into two daughter cells, one daughter cell is a clone of the stem cell and stays in the stem cell compartment and the other cell differentiates and leaves the stem cell compartment. Asymmetric cell division assures that the number of cells in the stem cell compartment, N_S , stays constant as new cells leave it, hence

$$N_{C,0}(t + \Delta t) = N_{C,0}(t) = N_S \tag{3}$$

for all t . $M_{C,0}(t)$ cells leave the stem cell compartment at each time step and enter the next compartment, i.e. the first compartment of the common precursors.

Common precursor compartments. The next compartment's size over time is

$$N_{C,1}(t + \Delta t) = N_{C,1}(t) - M_{C,1}(t) + M_{C,0}(t), \tag{4}$$

where $M_{C,1}(t)$ is the number of cells that leave it. The cells that leave this compartment divide symmetrically: Cells that divide themselves turn into two daughter cells and each of them differentiates and flows on to the next compartment. Contrary to previous work [13] we do not take into account potential self-renewal. With this simplification the inflow is $2M_{C,k-1}$ and

$$N_{C,k}(t + \Delta t) = N_{C,k}(t) - M_{C,k}(t) + 2M_{C,k-1}(t) \tag{5}$$

for all subsequent compartments $1 < k \leq K_C$ (see Fig 1). As there is no self-renewal in any of the non-stem-cell compartments the number of cell divisions a cell has gone through since it left the stem-cell compartment automatically determines its stage of maturation.

Committed precursor compartments and blood stream. What is the number of cells in the first compartment of each lineage? We assume that cells at the last common precursor stage randomly decide to go into one of the three blood cell lineages with probabilities $q_E(c(t))$, $q_G(c(t))$, and $q_T(c(t))$ that depend on the current growth factor concentrations $c(t)$:

$$\begin{pmatrix} M_{E,0}(t) \\ M_{G,0}(t) \\ M_{T,0}(t) \end{pmatrix} \sim \text{Multinomial} \left(M_{C,K_C}(t), \begin{pmatrix} q_E(c(t)) \\ q_G(c(t)) \\ q_T(c(t)) \end{pmatrix} \right) \tag{6}$$

for all t . Cells that leave the last compartment of the common precursors divide symmetrically, as do all subsequent compartments, and hence the outflow from the previous stage is always multiplied by two:

$$N_{x,k}(t + \Delta t) = N_{x,k}(t) - M_{x,k}(t) + 2M_{E,k-1}(t) \tag{7}$$

for $x \in \{E, G, T\}$ and $0 < k < K_x$. The last stage of each lineage models the blood stream, i.e. the number of erythrocytes N_{E,K_E} , granulocytes N_{G,K_G} , and thrombocytes N_{T,K_T} . Thrombocytes in the blood stream are not cells, instead platelets are generated from megakaryocytes and each megakaryocyte nuclear unit gives rise to many platelets. Hence, in order to be able to accommodate this fact we multiply the outflow of the previous to last stage with an additional factor m_x , therefore

$$N_{x,K_x}(t + \Delta t) = N_{x,K_x}(t) - M_{x,K_x}(t) + 2M_{x,K_x-1}(t) \cdot m_x \tag{8}$$

for $x \in \{E, G, T\}$. For the erythrocyte and granulocyte lineages m_E and m_G will be set to 1 whereas for the thrombocyte lineage m_T will be the number of platelets in a megakaryocyte nuclear unit.

Proliferation and death rates. It is known that within the bone marrow the proliferation rates increase with further differentiation (i.e. k). Stem cells divide only about once a year and the last committed precursors can divide several times a day. Dingli et al. [13] make the assumption that all rates are constant over time and that the ratio of rates $r = r_{x,k}/r_{x,k-1}$ at subsequent stages is constant, too. Therefore, $r_{x,k}$ is $r^k r_{x,0}$ for $1 < k \leq K_x$ and proliferation rates increase exponentially with differentiation stage k .

Here, we will make a similar but different assumption. The division rates at each stage should be controlled by the concentration of relevant growth factors (e.g. SCF, GM-CSF, Epo, G-CSF, TPO, etc.). We make the simplifying assumption that there are just three *effective* growth factors, one for each lineage. These could be Epo for the erythrocyte, G-CSF for the granulocyte, and TPO for the thrombocyte lineage—but more likely there is a combination of several factors in each lineage.

Each of the growth factors for the three lineages $x \in \{E, G, T\}$ has a certain concentration at each time point. Instead of using the concentrations for each growth factor directly it will be more convenient to parameterize the model with the logarithm of the concentrations $c_x(t)$. Let us again use the shorthand $c(t) = \{c_E(t), c_G(t), c_T(t)\}$ for the collection of all three concentrations. The division rates of a cell will depend on the probability that growth factors are bound to its receptors—which in turn depends on the growth factor concentrations. Note that the rates for each cell type can potentially depend on the concentrations of all three growth factors. In fact, we make the assumption that all common precursors (including the stem cells) have receptors for each of the three growth factors and in order to be able to divide, common precursors have to have at least one bound growth factor. In contrast, the committed precursors for the three lineages are only sensitive to their specific growth factor.

Let $s_{x,j}$ be the sensitivity for the specific growth factor of each lineage $x \in \{E, G, T\}$ at stage j where $0 \leq j < K_C + K_x$ (i.e. looking at Fig 1 the stem cells have index $j = 0$ and the first compartment of each committed lineage has index $j = K_C + 1$). The indices $0 \leq j \leq K_C$ capture the sensitivity of the common precursors to each of the three growth factors. For example, $s_{E,0}$ is the sensitivity of the stem cells to the growth factor of the erythrocyte lineage and s_{E,K_C} is the sensitivity of the last common precursor. The indices $K_C < j < K_C + K_x$ capture the sensitivities of the committed precursors. As cells at stage $K_C + K_E$ are the mature erythrocytes in the blood stream, the last stage that is sensitive to the E growth factor is $K_C + K_E - 1$. We assume that $s_{x,j}$ is a simple linear function of j , i.e. $s_{x,j} = a_x j + b$ for $0 \leq j < K_C + K_x$ with positive slope a_x and constant offset b . This is the key restriction in our model that replaces the constant ratio assumption of Dingli et al. [13]. Given these sensitivities the simplest model for the probability that a ligand is bound to a receptor is given by a logistic function:

$$p_{x,j}(c) = \frac{1}{1 + e^{-a_x j - b - c_x}} \quad \text{for } 0 \leq j < K_C + K_x, \tag{9}$$

where with a slight abuse of notation $c = c(t)$ is the collection of the logarithms of the three ligand concentrations and c_x pulls out the relevant concentration from c . If the diffusion of the growth factors in the bone marrow is much faster than the division rate of the cells, the binding of growth factors can be considered to be in steady state relative to the hematopoietic system. Under this assumption the equation for $p_{x,j}(c)$ can be derived from the principles of statistical mechanics (see e.g. the discussion of Hill functions in [18]).

Panel (A) in Fig 2 shows the binding probabilities for the three growth factors as a function of the stage j for all $c_x = 0$ (and a_x, b, K_C and K_x set to realistic values). If the concentrations increase the curves will shift to the left and the binding probabilities increase at all stages. For a decrease in concentrations the curves shift to the right and binding probabilities decrease.

We assume there is a maximum division rate r_{\max} that is determined by the cell cycle. This maximum rate is the same for all cell types. The division rate of each cell type reduces to the fraction of cells that have growth factors bound. For the common precursors we assume that they have three receptors, one for each growth factor of the respective lineage. We further assume they can divide if any of the three receptors have a growth factor bound to them. The probability that this happens at the common precursor stage k is the same as one minus the probability that none of the growth factors are bound and hence the rate at which these cells divide is

$$r_{C,k}(c) = r_{\max} \cdot (1 - (1 - p_{E,k}(c))(1 - p_{G,k}(c))(1 - p_{T,k}(c))) \tag{10}$$

for $0 \leq k \leq K_C$. The committed progenitors of each lineage only have a receptor for their respective growth factor and their proliferation rates are accordingly

$$r_{x,k}(c) = r_{\max} \cdot p_{x,K_C+k}(c) \tag{11}$$

for $x \in \{E, G, T\}$ and $1 \leq k < K_x$. The last stages K_E, K_G , and K_T of the three lineages are the erythrocytes, granulocytes, and thrombocytes in the blood stream. The rates at which each of them leaves the blood stream is assumed to be constant and is only determined by their respective death rates, i.e for each $x \in \{E, G, T\}$

$$r_{x,K_x}(c) = r_x, \tag{12}$$

and is therefore independent of the growth factor concentrations c . Note that the only way for a cell to die is to become a fully mature blood cell and then be cleared out of the blood stream after having served its purpose. Of course, in a more realistic model premature cells should also die sometimes. As it is hard to experimentally estimate the death rates for the various stages we have decided to ignore this possibility.

Panel (B) of Fig 2 gives a concrete numerical example of how the rates change for the four cell types as a function of the stage and again with all $c_x = 0$ (with r_{\max} and all r_x set to realistic values). There is a pronounced jump at K_C as the common progenitors have receptors for all three growth factors and the last common progenitor at K_C is therefore more active than each committed progenitor at the next stage. The rates drop to the death rates at the last stage of each lineage as these are not dependent on the binding probabilities (panel A). For the lower stages the rates look linear in log space and hence increase exponentially, as in the model of Dingli et al. [13]. In contrast to their model the rates in our model saturate at r_{\max} .

From last common precursor to the three lineages. We will use the binding probabilities from Eq (9) to set the probabilities $q_E(c), q_G(c)$ and $g_T(c)$ that common precursor cells enter the three lineages (see Eq 6). Let us look at the erythrocyte lineage first. We will assume that if a cell at the last stage of the common precursors has only an E growth factor bound and not any of the other two then the cell will enter the erythrocyte lineage. If it has an E growth factor and exactly one of the other two, the probability to enter the E lineage is 1/2. If all growth factors

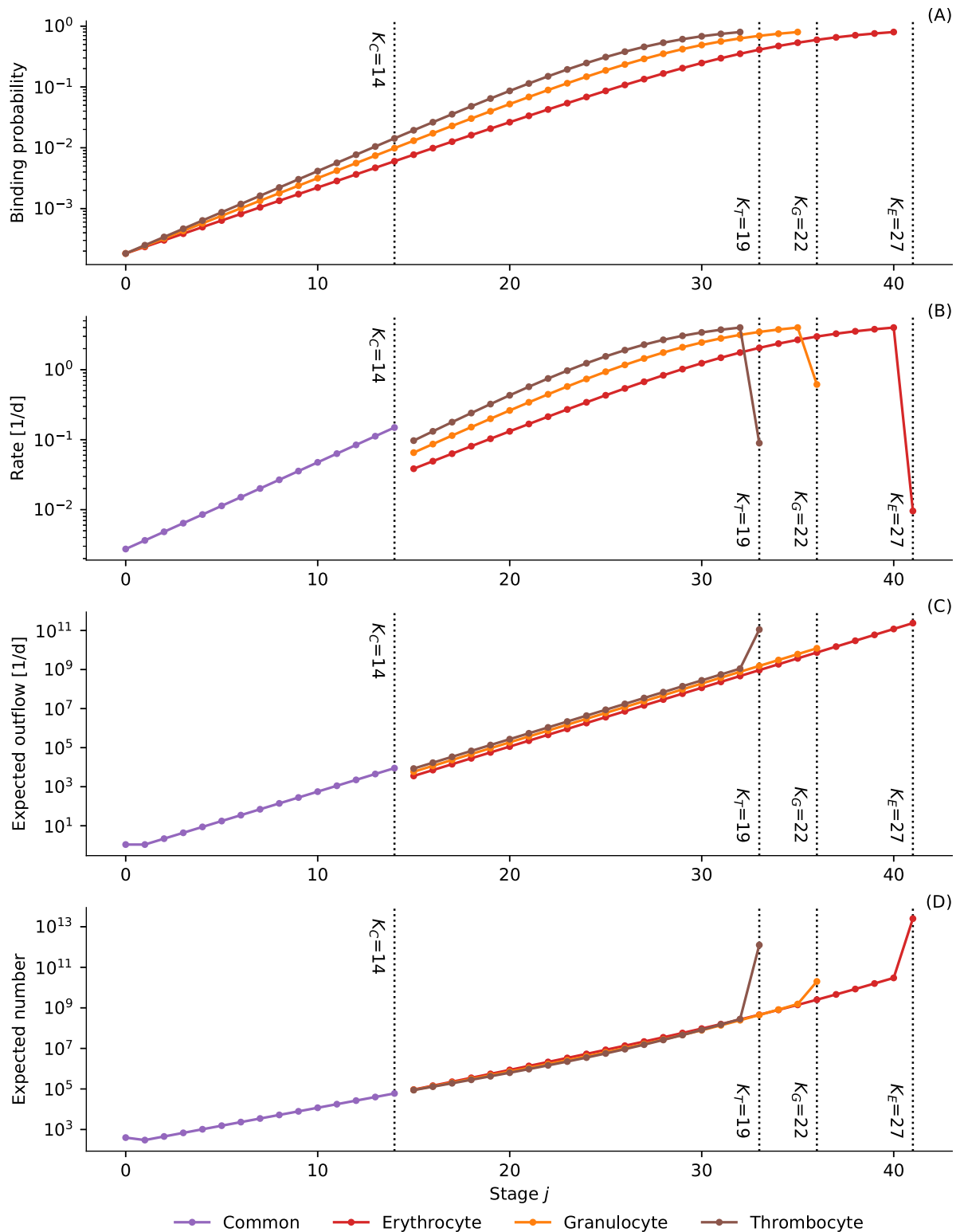


Fig 2. Model stages. Panel (A) shows binding probabilities for the erythrocyte, granulocyte, and thrombocyte growth factors as a function of the differentiation stage. Panel (B) shows the corresponding proliferation rates, except for the last stages of each lineage that model the blood stream. For those the rates correspond to the death rates of erythrocytes, granulocytes, and thrombocytes. Panels (C) and (D) depict the expected outflow and number of cells at each stage. All curves are for the growth factor concentrations at their target values (all $c_x = 0$) and with all parameters set to realistic values.

<https://doi.org/10.1371/journal.pone.0204393.g002>

are bound the probability is 1/3. Let

$$\begin{aligned}
 q'_E(c) &= p_{E,K_C}(c)(1 - p_{G,K_C}(c))(1 - p_{T,K_C}(c)) \\
 &+ \frac{1}{2}p_{E,K_C}(c)p_{G,K_C}(c)(1 - p_{T,K_C}(c)) + \frac{1}{2}p_{E,K_C}(c)(1 - p_{G,K_C}(c))p_{T,K_C}(c) \\
 &+ \frac{1}{3}p_{E,K_C}(c)p_{G,K_C}(c)p_{T,K_C}(c)
 \end{aligned}$$

and similarly for q'_{G,K_C} and q'_{T,K_C} . With these three values we can compute the probability that a cell that divides at the last common precursor stage enters lineage $x \in \{E, G, T\}$ as

$$q_x(c) = \frac{q'_x(c)}{q'_E(c) + q'_G(c) + q'_T(c)}. \tag{13}$$

Feedback mechanism. The last missing component for the model is the specification of the feedback mechanism. We assume that feedback regulates the growth factor concentration for each lineage $x \in \{E, G, T\}$ separately. If the number $N_{x,K_x}(t)$ of cells or platelets in the blood is lower than the target number N_x the concentration goes up, otherwise it goes down, i.e.

$$c_x(t) = \frac{\delta}{N_x} (N_x - N_{x,K_x}(t)) \tag{14}$$

where δ/N_x is the gain of the feedback mechanism that, for simplicity, we assume to be the same for all three growth factors relative to the respective target number N_x . Hence, the feedback mechanism tries to keep all $c_x(t)$ at zero and therefore $N_{x,K_x}(t)$ close to N_x . While the feedback signal is a linear function of the deviation, its effect on the binding probabilities (Eq 9)—and therefore the proliferation rates at each stage (Eq 11)—is non-linear. Also, the feedback mechanism is deterministic and will reduce the overall noise in the system. In fact, we will see below (subsection on *noise scaling*) that the system behaves almost deterministically and extend it in a way that better reflects the noise in biological systems.

Results

Now that we have presented our model in full detail, the *Results* section will first give a theoretical analysis of the steady-state behavior of the model. Then this analysis will be used to set the model's parameters to physiologically sensible values. Finally, we extend the model slightly to allow for simulating the effects of single mutations on leukemogenesis.

Conditional steady state analysis

The proliferation rates at all stages and the probability of cells to enter each lineage depend on the binding probabilities in Eq (9). These, in turn, are completely determined by the growth factor concentrations $c_x(t)$ in the bone marrow. The concentrations for the three growth factors therefore control the overall output of cells for all three lineages and are themselves controlled by the feedback mechanism that keeps the number of erythrocytes, granulocytes, and thrombocytes close to their target values (Eq 14). To better understand this system we will analyze its steady state behavior for given, fixed concentrations $c_x(t) = c_x$. For the following derivations we will again collect these three numbers in one variable $c = \{c_E, c_G, c_T\}$.

In steady state the expected number of cells at each state is constant and the expected inflow equals the expected outflow on all stages. For the common precursors this means that Eq (4)

can be simplified to

$$E[M_{C,1} | c] = E[M_{C,0} | c]$$

where we have dropped the dependence on t since all time points behave in the same way in steady state. By the same reasoning Eq (5) can be simplified to

$$E[M_{C,k} | c] = 2E[M_{C,k-1} | c] \quad \text{for } 1 < k \leq K_C.$$

From these two equations it follows that

$$E[M_{C,k} | c] = 2^{k-1}E[M_{C,0} | c] \quad \text{for } 1 < k \leq K_C.$$

For the three downstream lineages $x \in \{E, G, T\}$ the same reasoning can be applied to Eqs (7) and (8) and results in

$$\begin{aligned} E[M_{x,k} | c] &= 2^k E[M_{x,0} | c] && \text{for } 1 < k < K_x \\ E[M_{x,k} | c] &= 2^k E[M_{x,0} | c] \cdot m_x && \text{for } k = K_x \end{aligned}$$

where again the last stage of each lineage K_x needs to be treated differently because there is the additional factor m_x to accommodate the fact that megakaryocytes burst into m_T platelets.

The expected inflow for all three lineages $x \in \{E, G, T\}$ depends on the expected outflow of the last common progenitor stage and the probability $q_x(c)$ that a cell enters lineage x (Eq 6) is

$$E[M_{x,0} | c] = E[M_{C,K_C} | c] \cdot q_x(c)$$

because c is given and therefore $q_x(c)$ is not random. Since $E[M_{C,K_C} | c] = 2^{K_C-1}E[M_{C,0} | c]$ the preceding equations can be summarized as

$$E[M_{C,k} | c] = E[M_{C,0} | c] \cdot 2^{k-1} \quad \text{for } 1 \leq k \leq K_C \tag{15}$$

$$E[M_{x,k} | c] = E[M_{C,0} | c] \cdot 2^{K_C+k-1} \cdot q_x(c) \quad \text{for } 1 \leq k < K_x \tag{16}$$

$$E[M_{x,k} | c] = E[M_{C,0} | c] \cdot 2^{K_C+k-1} \cdot q_x(c) \cdot m_x \quad \text{for } k = K_x \tag{17}$$

where $x \in \{E, G, T\}$. This shows that the expected outflow of all stages *only* depends on the expected outflow of the stem cell compartment $E[M_{C,0}|c]$. Hence, using these equations the expected outflow for all compartments can be computed by starting with the expected outflow of the stem cell compartment

$$E[M_{C,0} | c] = E[N_{C,0} | c]r_{C,0}(c) \cdot \Delta t = N_s r_{C,0}(c) \cdot \Delta t$$

where we have used that $M_{C,0}$ is binomially distributed (Eq 2) and that the number of stem cells is constant (Eq 3). Panel (C) in Fig 2 shows the expected outflow of cells per day for each stage for realistic parameter values and all $c_x = 0$. The constant slope in log space is due to the doubling at each stage for all cell types—except for the stem cell compartment and the last stage of the thrombocyte lineage. In the former case there is asymmetric cell division and therefore the outflow of the stem cell compartment and the next compartment are the same. In the latter case each megakaryocyte results in many more than just two platelets ($m_T \gg 1$ whereas $m_E = m_G = 1$).

From the expected outflow we can compute the expected number of cells at each stage. For all the left hand sides of Eqs (15), (16) and (17)

$$E[M_{x,k} | c] = E[N_{x,k} | c]r_{x,k}(c) \cdot \Delta t$$

because, again, the $M_{x,k}$ are binomially distributed (Eq 2). Plugging the last two equations back into Eqs (15), (16) and (17) we find that the expected number of cells at each stage is

$$E[N_{C,k} | c] = \frac{N_S r_{C,0}(c)}{r_{C,k}(c)} \cdot 2^{k-1} \quad \text{for } 1 \leq k \leq K_C \tag{18}$$

$$E[N_{x,k} | c] = \frac{N_S r_{C,0}(c)}{r_{x,k}(c)} \cdot 2^{K_C+k-1} \cdot q_x(c) \quad \text{for } 1 \leq k < K_x \tag{19}$$

$$E[N_{x,k} | c] = \frac{N_S r_{C,0}(c)}{r_{x,k}(c)} \cdot 2^{K_C+k-1} \cdot q_x(c) \cdot m_x \quad \text{for } k = K_x. \tag{20}$$

Panel (D) in Fig 2 shows the number of expected cells for each stage in the hematopoietic system for realistic parameter values and all $c_x = 0$. The expected number drops a little in the first compartment after the stem cell compartment because the rate increases (panel B) but the outflow does not (panel C). The big jumps on entering the blood stream at $K_C + K_x$ are due to the fact that the death rates r_x are much lower than the proliferation rates of the last progenitors (panel B).

The blood stream, i.e. the last stage K_x of each lineage, is also special because it produces the feedback to all the other stages of the hematopoietic system. The concentration of each growth factor c_x is a deterministic and one-to-one function of the number of cells in each last compartment N_{x,K_x} (Eq 14). Hence, if c_x is known so is N_{x,K_x} . In particular, if $c_x = 0$

$$E[N_{x,K_x} | c = 0] = N_x$$

where we used the abbreviation $c = 0$ to mean all $c_x = 0$ in $c = \{c_E, c_G, c_T\}$. Using the same abbreviation let $r_S = r_{C,0}(0)$ be the proliferation rate of the stem cells for $c = 0$, i.e. this is the rate that the feedback mechanism aims for (Eq 14). Remember that $r_x = r_{x,K_x}(0)$ are the death rates of each lineage in the blood stream independent of the growth factor concentrations (Eq 12). Plugging all these values into Eq (20), the conditional expected number of erythrocytes, granulocytes, and thrombocytes at the final stages hence have to fulfill the following constraints:

$$N_x = \frac{N_S r_S}{r_x} \cdot 2^{K_C+K_x-1} \cdot q_x \cdot m_x \tag{21}$$

for $x \in \{E, G, T\}$ and with $q_x = q_x(0)$.

Furthermore, let $N_C = E[N_{C,K_C} | c = 0]$ and $r_C = r_{C,K_C}(0)$ be shorthands for the respective values of the last common progenitor compartment. Plugging these into Eq (18) for $k = K_C$ gives one more constraint, namely

$$N_C = \frac{N_S r_S}{r_C} \cdot 2^{K_C-1}. \tag{22}$$

These constraints can be used to set realistic values for the free parameters of the model. Importantly, the above steady state analysis is conditional on all $c_x = 0$ and hence ignores the feedback mechanism (Eq 14). The unconditional steady state is unfortunately more complicated to analyze, also because the feedback signal has a non-linear influence on the system.

However, the hope is that setting parameters based on the conditional steady state analysis will still lead to useful values—which is indeed the case as shown next.

Rough parameter estimates

The free parameters of the model are:

- the number of stem cells N_S
- the target number of erythrocytes, granulocytes, and thrombocytes in the blood stream, N_E , N_G , and N_T
- their death rates r_E , r_G , and r_T
- the number m_T of platelets in each megakaryocyte
- the maximum proliferation rate of cells r_{\max}
- the sensitivity offset parameter b
- the increase in growth factor sensitivity over stages for each lineage a_E , a_G , and a_T
- the number of stages for each lineage K_E , K_G , K_T
- and for the common precursors K_C
- and, finally, the feedback gain δ .

We will use clinical observations and current best estimates to set these free parameters. First, for the number of stem cell N_S we use the same estimate as Dingli et al. [13], namely that there are roughly 400 hematopoietic stem cells that are active at any one time [19]. For healthy adults

$$\begin{aligned} N_E &\approx 5.0 \text{ million}/\mu\text{L} \cdot 5L \approx 25.0 \text{ trillion erythrocytes} \\ N_G &\approx 4000/\mu\text{L} \cdot 5L \approx 20.0 \text{ billion granulocytes} \\ N_T &\approx 250000/\mu\text{L} \cdot 5L \approx 1.2 \text{ trillion thrombocytes} \end{aligned}$$

are realistic numbers that can be obtained from clinical norms and assuming 5 liters of blood in the body. Erythrocytes are estimated to die after 110 days on average ($1/r_E$). For granulocytes the available estimates are highly variable. The latest estimates using in vivo labeling are 5 days, considerably higher than previous ex-vivo estimates [20]. However, these measurements were just for one type of granulocyte and other estimates suggest survival times of only 8 hours. We therefore decided to use a value in between, namely 2 days ($1/r_G$). Classic data on thrombocyto-kinetics estimate that thrombocytes leave the blood stream after 10 days ($1/r_T$). One complication with thrombocytes is that the last committed precursor in the bone marrow, the megakaryocytes, produce about a thousand platelets. Another complication is that megakaryocytes undergo endomitosis, i.e. each cell cycle doubles the chromosomes and nuclear units within the cell but there is no cell division. In the model we will not distinguish between normal mitosis and endomitosis and hence m_T is the number of thrombocytes in each nuclear unit of a megakaryocyte. This number is estimated to be around 50 [21]. In contrast, $m_E = m_G = 1$ as the last erythrocyte and granulocyte precursors divide normally. Finally, the cell cycle takes a certain time and therefore cells cannot divide at arbitrary rates. A realistic number for the maximum number of cell divisions per day, r_{\max} is 5 [13].

The remaining parameters are more difficult to set to reasonable values as their relationship to more easily observable quantities can be non-linear and parameters interact, too. For

example, from Eq (9) it follows that the offset parameter b and the slopes a_x are related to the log odds of the binding probability at each stage j , i.e.

$$a_x j + b + c_x = \log\left(\frac{p_{x,j}(c)}{1 - p_{x,j}(c)}\right) \tag{23}$$

for $x \in \{E, G, T\}$. For $j = 0$ we can, however, solve this equation for b . If all $c_x = 0$, the three probabilities in the stem cell compartment, $j = 0$, are equal, i.e. $p_{\text{stem}} = p_{E,0}(0) = p_{G,0}(0) = p_{T,0}(0)$. This is the situation when erythrocytes, granulocytes, and thrombocytes are all at their target values (Eq 14). With this assumption

$$b = \log\left(\frac{p_{\text{stem}}}{1 - p_{\text{stem}}}\right).$$

and using Eq 11 and again using the shorthand $r_s = r_{C,0}(0)$ we find that

$$\begin{aligned} r_s &= r_{\text{max}} \cdot (1 - (1 - p_{\text{stem}})^3) \\ (1 - p_{\text{stem}})^3 &= \left(1 - \frac{r_s}{r_{\text{max}}}\right) \\ p_{\text{stem}} &= 1 - \left(1 - \frac{r_s}{r_{\text{max}}}\right)^{\frac{1}{3}}. \end{aligned} \tag{24}$$

We already know r_{max} but what is the proliferation rate of the stem cells r_s ? The rate at which they divide is about once a year, i.e. 1/365. This is the same estimate as used by Dingli et al. [13] which is based on allometric scaling arguments [19]. It is in between other estimates that range from once every 25-50 weeks to 0.6 times per year [22, 23]. Anyway, with this number we can compute p_{stem} and we find that $b \approx -8.61$.

With b known, what are the slopes a_E , a_G , and a_T ? We assume that we know the rate at the previous to last stage of each lineage (remember the last stage are the mature blood cells), the output rate into the blood stream: $r_{\text{blood}} = r_{E,K_E-1} = r_{G,K_G-1} = r_{T,K_T-1} = 4$ times per day. This is chosen to be just below the maximum rate but leaving a little room to increase the rate for the feedback mechanism. With this assumption we immediately get from Eq (11) that $p_{\text{blood}} = p_{x,K_C+K_x-1} = r_{x,K_x-1}/r_{\text{max}}$. Which allows us to solve Eq (23) for a_x :

$$a_x = \frac{\log\left(\frac{p_{\text{blood}}}{1 - p_{\text{blood}}}\right) - b}{K_C + K_x - 1}$$

for $x \in \{E, G, T\}$. Unfortunately we can only solve this if we know the number of stages K_E , K_G , K_T , and K_C . From Eq (21) we get that

$$K_x = 1 - K_C + \log_2\left(\frac{N_x r_x}{N_s r_s}\right) - \log_2(q_x) - \log_2(m_x) \tag{25}$$

for $x \in \{E, G, T\}$. Hence, we can only compute K_x if we know all a_x and K_C since the q_x depend on those. But remember that a_x in turn depends on all K_x , including K_C . Assuming that at least K_C is known, say $K_C \approx 14$, these mutual dependencies suggest an iterative scheme to update each set of equations in turn. This scheme converges extremely quickly. If applied we find that $a_E \approx 0.25$, $a_G \approx 0.29$, and $a_T \approx 0.31$. More importantly, we get estimates for the number of stages in each lineage, namely

$$K_E \approx 27 \quad K_G \approx 22 \quad K_T \approx 19$$

that we rounded to the closest integer. Hence, in order to become an erythrocyte a stem cell has to go through $K_E + K_C \approx 41$ cell divisions. This number is considerably higher than previous estimates [13] (in their Eq (1)—that is analogous to our Eq (25) with $q_x = 1$ and $m_x = 1$ —they do not have r_S).

Due to the rounding of K_x these estimates will not produce the exact desired expected out-flow $N_x r_x$ for the hematopoietic system that was used to fit the parameters. As we can only have an integer number of stages, the target numbers of erythrocytes, granulocytes, and thrombocytes are constrained by the integer powers in Eq (21). In order to make the N_x fit the desired values we adjust the r_x , i.e. the death rates of erythrocytes, granulocytes and thrombocytes according to Eq (21). The new adjusted estimates are

$$1/r_E \approx 104.41 \text{ days} \quad 1/r_G \approx 1.62 \text{ days} \quad 1/r_T \approx 11.20 \text{ days}$$

which are still realistic compared to the initial values of 110, 2, and 10 days.

These estimates depend on knowing K_C that we have seemingly arbitrarily set to 14. Based on mutations that affect several lineages it has been estimated that there are between 20 and 100 thousand CFU-GEMM cells [13], i.e. last common progenitors. Say we want to choose K_C such that the expected number of cells at the last common precursor stage is approximately 60000. We can vary K_C , fit all other parameters given K_C and compute the resulting conditional expected number of last common precursors $N_C = E[N_{C,K_C} | c = 0]$ using Eq (22). We can then choose K_C such that $N_C \approx 60000$. Using this procedure results in $K_C \approx 14$, the number we have used above. The expected behavior of the model with all the parameters set as just described is shown in Fig 2. Recently, it has been suggested that the textbook view of many stages of pluripotent common progenitors in hematopoiesis may not be correct and that cells commit very early to one of the three lineages [24]. If this was true K_C would have to be chosen much smaller, perhaps even set to zero. In the following we will, however, stick to the textbook view.

Feedback and stem cell transplants. Finally, we set the gain of the feedback mechanism δ . If the granulocyte count drops to zero, as it can for example happen after radiation- or chemo-therapy, the stem cells should definitely proliferate more than at their steady state rate r_S , but probably also not at the same maximum rate r_{max} that cells towards the end of the system work at (see Fig 2, panel B). Let r'_S be the rate that results in the case when all $c_x = \delta$ (see Eq 14). There is a one-to-one correspondence between r'_S and δ that can be obtained from Eqs (23) and (24) when substituting r'_S for r_S and δ for c_x :

$$\delta = \log\left(\frac{p'_{stem}}{1 - p'_{stem}}\right) - b \quad p'_{stem} = 1 - \left(1 - \frac{r'_S}{r_{max}}\right)^{\frac{1}{3}}$$

Setting r'_S is more intuitive than setting δ directly. As the amplification of the output of the stem cell compartment is so massive (Eq 17) even a small increase from r_S to r'_S can lead to a big change in the blood counts. Also, normal fluctuations around the steady state should not lead to big fluctuations of the proliferation rate of the stem cells. Still, even with these constraints it is challenging to set r'_S without detailed data on the dynamics of the system. We will set this parameter by simulating a stem cell transplant and making sure that the time course of the simulated parameters roughly matches clinical observations. In this simulation, and most simulations to be reported below, we have set $\Delta t = 1/r_{max}/5$. We have also simulated most results of this paper with a Δt twenty times smaller and do not find significant deviations in the results. For the few simulations that simulated several years we set $\Delta t = 1/r_{max}$.

The left side panels of Fig 3 depict the erythrocyte, granulocyte, and thrombocyte counts after a stem cell transplant that happened at $t = 0$. To simulate the stem cell transplant we first

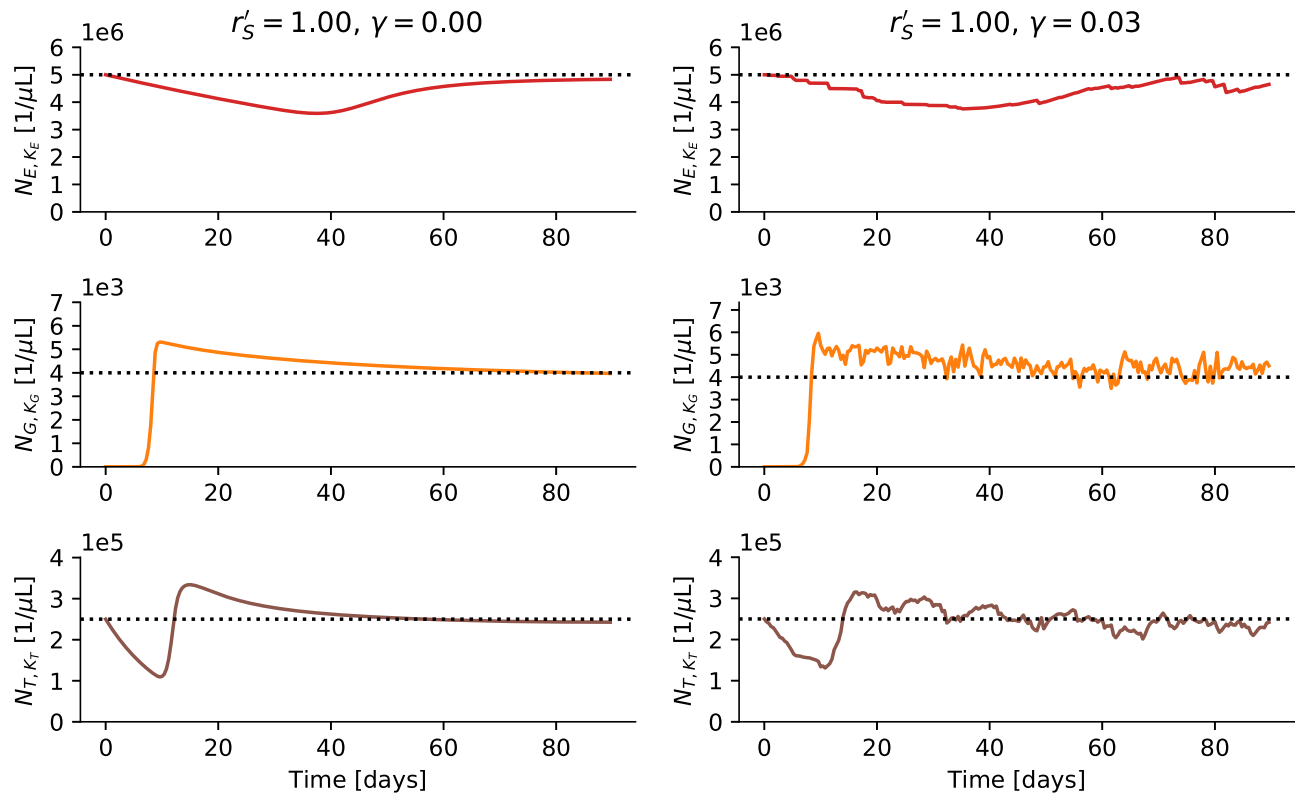


Fig 3. Simulation of stem cell transplant. The left side panels depict erythrocyte, granulocyte, and thrombocyte counts for a simulation of a stem cell transplant (for details see main text). The right side panels show the same with increased noise scaling.

<https://doi.org/10.1371/journal.pone.0204393.g003>

simulate the effect of chemo-therapy or radiation-therapy. After a couple of weeks of therapy we assume that all blood production came to a halt and the reserves in the bone marrow have been used up, i.e. there are no more cells at any of the stages of the model, except the blood stream. The erythrocytes and thrombocytes are assumed to be at normal levels because they have been transfused to stabilize the patient. Granulocytes, in contrast, cannot be transfused easily and hence because of their short half-life we assume they have been completely depleted (patients are treated with antibiotics to prevent infections). Then, at $t = 0$, a short time after the transplantation all stem cells are assumed to be back to their normal functioning and start replenishing the bone marrow. It then takes a while until the first fresh cells arrive in the blood stream. There is an overshoot in the granulocytes and thrombocytes and the system takes a while to return to the steady state. Erythrocytes respond a lot more slowly. The time until the system starts responding (the granulocyte counts start going up) shortens with an increase in r'_s but this also results in a bigger overshoot and a longer time to resettle. We have set $r'_s = 1.00$ by eye such that it takes less than two weeks until granulocytes and thrombocytes are back up but the overshoot does not become unrealistically large.

Comparison to clinical data. The time scale of less than two weeks is consistent with clinical experience simply because we have set the parameters accordingly. But do the details of the dynamics that our model shows fit clinical observations, too? This is not easy to answer. Allogeneic transplantation has the complication of graft versus host disease, hence it is simpler to consider autologous transplantation only. As a first validation of the model, we have therefore looked up clinical records of 24 lymphoma patients who received autologous

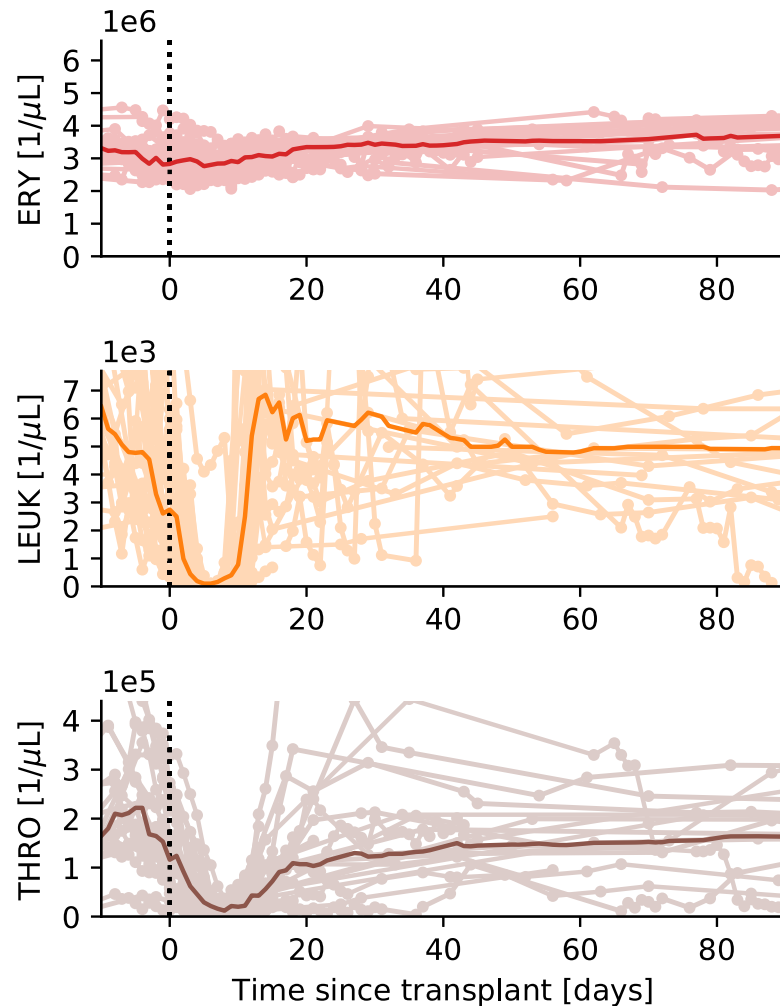


Fig 4. Clinical data from autologous stem cell transplants. The erythrocyte, leukocyte, and thrombocyte counts for 24 lymphoma patients around the time of their stem cell transplant are shown. The faint colors depict the individual patients with dots denoting actual measurements. The bold colors depict the median time-courses of linear interpolations of the single-patient data.

<https://doi.org/10.1371/journal.pone.0204393.g004>

hematopoietic stem cell transplantation after myeloablative conditioning. The time courses of their blood counts can be seen in Fig 4 along with the median time-course.

Consider the erythrocytes first. As expected, the model and the data show a very slow time-course with relatively small changes over the 90 days after the transplant. The main difference is that the patients are anemic. As we calibrated the model with data from healthy subjects it is not surprising that the absolute numbers are different.

As for the leukocytes: The simulation only models the granulocytes as the most frequent leukocytes and therefore the numbers in Figs 3 and 4 are not directly comparable but the median time-course of the overshoot and the slow resettling to the steady state are remarkably similar—even though we only gathered these data after having fixed the parameters.

The thrombocyte counts show a very different median time-course than expected. In particular, there is no overshoot in the clinical data. It could be that we simply need to change the gain parameter for the thrombocyte feedback loop separately from the other gains (Eq 14). However, in contrast to leukocytes, thrombocytes can be transfused and patients do receive

transfusions during their treatment. Our simulation assumed therefore that patients go into the transplantation with a normal thrombocyte count. This is clearly not the case in the data. Patients receive transfusions at varying time-points during their treatments and these transfusions have a big effect on the thrombocyte counts. Also, patients vary in the time that passed between the myeloablative conditioning and the stem cell transplant. There might be an increased demand of thrombocytes during the transplant, too. Given all these uncontrolled factors, the median time-course of the patient data is not easily comparable to the model simulation for the thrombocyte counts. In general, given the large differences between single patients it seems desirable for future work to try to model single patients and not just prototypical behavior.

Noise scaling. A striking feature of the simulation depicted in the left panels of Fig 3 is that the curves are extremely smooth and the system seems to behave almost deterministically, despite the fact that random choices are made at every stage of the system. Apparently the numbers of cells are so big, the difference between inflow and outflow so small, and the feedback mechanism so effective that there is hardly any significant variation in cell counts. In fact, when running the simulation for a year in steady state the absolute average deviation from the target values is at most 0.026%. Such low variance in the blood counts is, of course, unrealistic. In reality there will be many more sources for variability in the data, e.g. additional noise in the growth factor concentrations due to infections or temporarily increased oxygen demand. The simulation should look more like the right panels in Fig 3. As the main source of variance in the model are the binomial random variables (Eq 2) one simple way to inject more noise into the system is to change the binomial to a beta-binomial to allow for overdispersion. This can be achieved by changing Eq (1) to

$$\begin{aligned}
 r'_{x,k}(t) &\sim \text{Beta}\left(\frac{\alpha_{x,k}(t)}{\gamma}, \frac{\beta_{x,k}(t)}{\gamma}\right) \\
 \alpha_{x,k}(t) &= r_{x,k}(c(t)) \\
 \beta_{x,k}(t) &= 1 - \alpha_{x,k}(t)
 \end{aligned}$$

and in this way the expectation of $M_{x,k}(t)$ stays the same (Eq 2), hence keeping the conditional steady state analysis intact. The variance, however, increases with the noise scaling parameter γ . For $\gamma \rightarrow 0$ the original model is recovered. In the limit of $\Delta t \rightarrow 0$ a beta process is obtained instead of the Poisson process of the original model [25].

Unfortunately, the introduction of the noise scaling parameter γ leads to shifts in the unconditional steady-state values of the erythrocyte, granulocyte, and thrombocyte counts away from their target values N_E , N_G , and N_T (Eq 14). The left side panels in Fig 5 show the steady state behavior of the system for $\gamma = 0.03$. There are small shifts of the average counts compared to the desired target values (the dashed horizontal lines) but they are hardly noticeable. In the panel on the right the relationship between γ and the expected values of the steady state behavior is explored systematically. The noise scaling parameter γ is varied and the deviation of the average counts from their target values over a simulation of one year is shown. The deviations increase with the noise scaling parameter. This effect can be explained by the non-linear feedback that quickly increases the proliferation rate of the stem cells whenever the output is too low but cannot decrease the proliferation rate to the same degree whenever the output is too high—simply because the proliferation rate of the stem cells is already very low for the target values. In the following we have set $\gamma = 0.03$. This value was chosen by eye to give a reasonable range for the variances of the erythrocytes, granulocytes, and thrombocytes but results only in a relatively small average deviation from the target values.

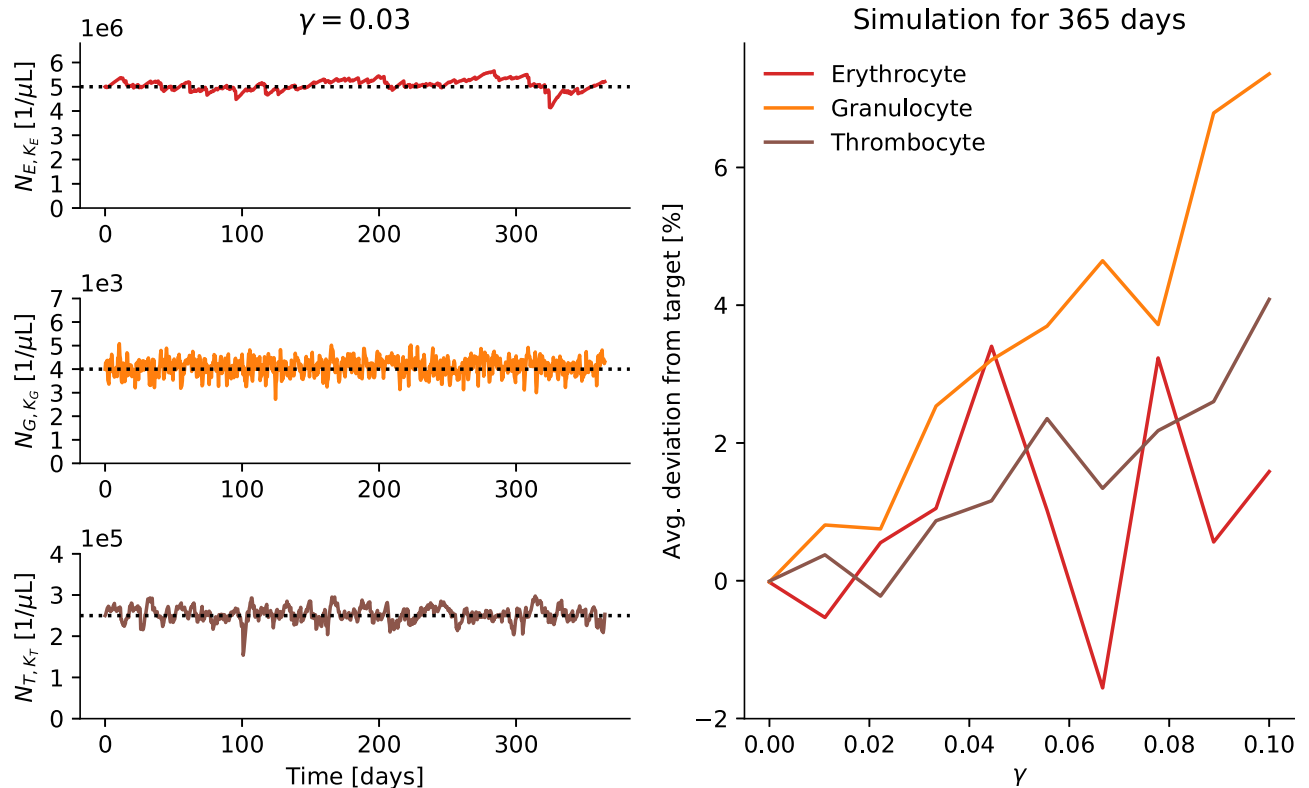


Fig 5. Deviation from target values. The panels on the left show the erythrocyte (N_{E,K_E}), granulocyte (N_{G,K_G}), and thrombocyte (N_{T,K_T}) counts for the hematopoietic system in steady state. The noise scaling parameter γ is varied in the panel on the right and the resulting average deviation of the counts from their target values is shown for one year-long run of the simulation.

<https://doi.org/10.1371/journal.pone.0204393.g005>

Simulation of myeloid leukemia

Ultimately, models for hematopoiesis should be useful for understanding clinical conditions. As a first step in this direction, we will extend the model to capture some aspects of acute myeloid leukemia (cf. [16]). We assume leukemia develops through mutations in crucial pathways, the so-called hallmarks of cancer [8, 9, 26]. Here, we will focus on two such pathways: The pathway that controls the replicative capacity of a cell and the pathway that controls its further differentiation. We extend the above model of hematopoiesis by allowing for two kinds of mutations, one in each pathway.

First, we allow for mutations that give cells unlimited replicative capacity. This requires the introduction of a limited replicative capacity for “normal” cells (which was ignored in [16]). Normal cells are limited in their replicative capacity by the Hayflick limit. We assume all cells that leave the stem cell compartment can replicate a fixed number of times, and this number is the same for all cells. This number has to be greater than the number of stages that the erythrocytes have to go through—otherwise no mature erythrocytes will be produced—but decreases with increasing age and for adults should therefore be lower than the absolute Hayflick limit of about 70. Here we have arbitrarily set it to 60. In the model we keep track of the remaining number of divisions for a cell by adding subcompartments to each stage accordingly: A compartment collects all cells at a certain stage with a certain remaining replicative capacity. For normal cells the remaining replicative capacity decreases as they differentiate further and cells that have exhausted their replicative capacity vanish from the simulation. Unlimited replicative

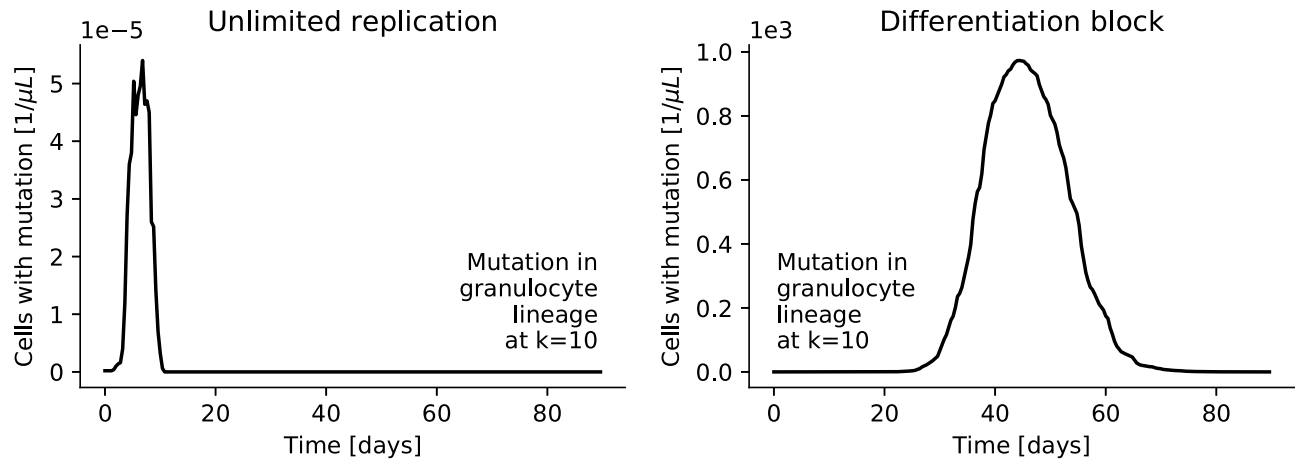


Fig 6. Single mutations. Mutations are introduced in the granulocyte lineage at stage $k = 10$ at time zero. In the left panel the mutation gives the mutated cell and all its daughter cells unlimited replicative capacity. The plot shows the number of cells with this mutation as a function of time (measured per microliter blood in the body). The clones quickly die out because they differentiate into mature granulocytes and are subsequently removed from the system. In the right panel the mutation blocks further differentiation but also these clones die out once they have exhausted their replicative capacity.

<https://doi.org/10.1371/journal.pone.0204393.g006>

capacity can now be modeled by simply not decreasing the replicative capacity of a cell. That is, the two children of a cell move on to the next differentiation stage without a decrease in their replicative capacity and so do all their children and all their children, and so on, giving rise to an exponential growth of cells with this particular mutation. For the left panel of Fig 6 such a mutation was introduced at the tenth stage of the granulocyte lineage. The number of clones with this mutation is shown as a function of time for one run of the simulation. Even though these cells have unlimited replicative capacity they quickly die out as the daughter cells eventually differentiate further into mature granulocytes that do not divide any more and are quickly removed from the blood stream.

Second, we allow cells to develop a complete differentiation block, i.e. they do not differentiate further. Instead all their daughter cells (and their children and so on) remain at the same stage of the hematopoietic system. A differentiation block leads to an exponential increase in cells at the stage where the block occurs. The proliferation rate is given by the proliferation rate of cells at this stage (see panel B in Fig 2). Hence, a mutation at an earlier stage will lead to a slower increase in clones than a mutation at a later stage. However, the exponential increase will eventually slow down and will be followed by a decrease of clones once the clones have exhausted their replicative capacity. This time course is depicted in the right panel of Fig 6. The peak number of cells depends on the stage where the mutation occurs. Later stages have already used up more of their replicative capacity and therefore cannot proliferate as much. In order to get a number of mutated cells in the bone marrow that is comparable in magnitude to the number of normal cells in the bone marrow during steady state the mutation has to occur at a very early stage.

Say, a differentiation block occurs at the common precursor stage $k = 8$, much earlier than in Fig 6. As cells at this early stage have not used up as much of their replicative capacity as later cells they can produce a much larger number of descendant cells. Fig 7 shows that a differentiation block that occurs early can produce a number of clones that produce problems for hematopoiesis. In order to make the model a bit more realistic, we have assumed that there is a hard limit to the number of cells that can be hosted by the bone marrow, either because of space limitations or other finite resources. Excess cells—mutated and normal cells—are

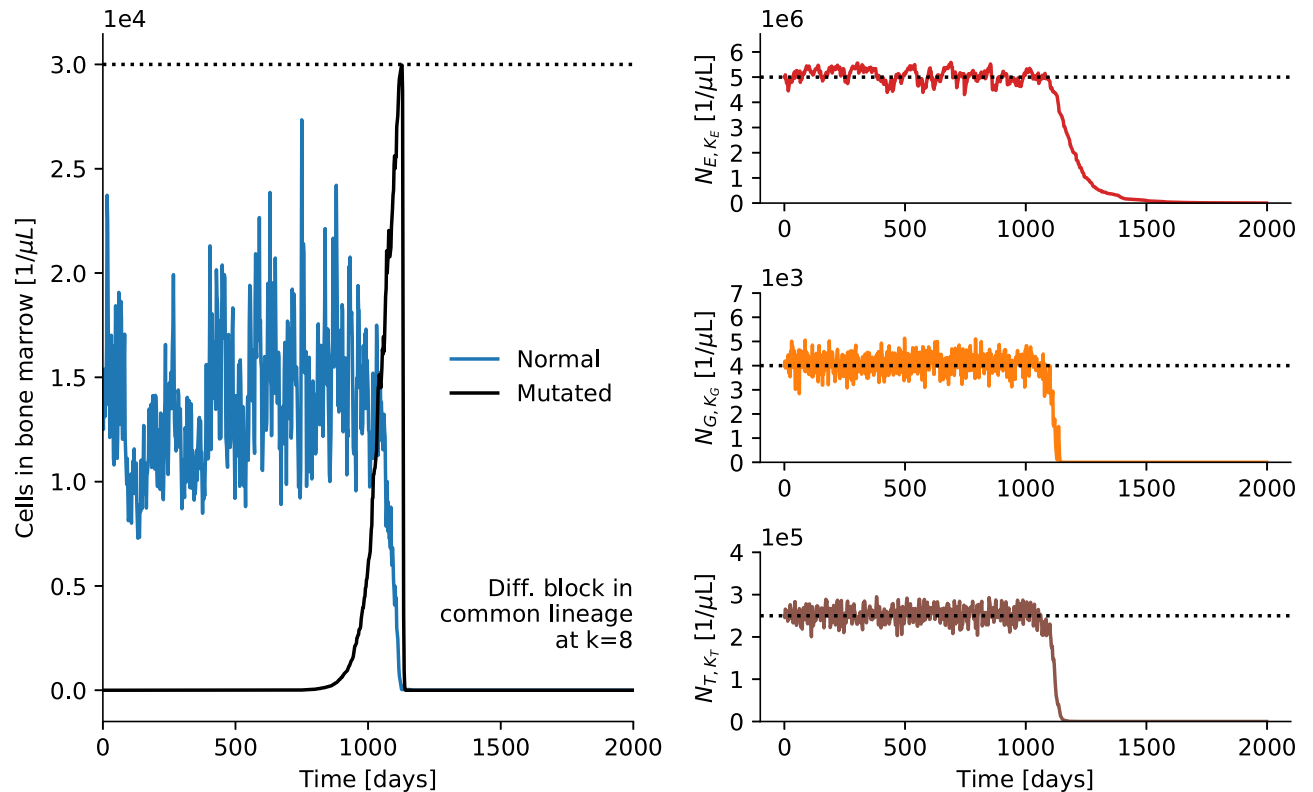


Fig 7. Early differentiation block. A differentiation block is introduced in the common precursor stage $k = 8$ at time zero. The blue plot in the left panel shows the number of normal cells in the bone marrow (per microliter blood in the body). After a nascent phase with slow growth the number of clones, shown in black, explodes. As in the right panel of Fig 6 the number of clones here is bound by a limited replicative capacity and comes down again. However, here we assume in addition that there is a limit to the number of cells that can be hosted by the bone marrow (dashed horizontal line) and excess cells are pushed into the blood stream. The panels on the right show the corresponding development of erythrocyte, granulocyte, and thrombocyte counts in the blood. After the bone marrow has been taken over completely by the mutated cells there is no more production of mature blood cells and hematopoiesis breaks down completely.

<https://doi.org/10.1371/journal.pone.0204393.g007>

assumed to be pushed out of the bone marrow and into the blood stream at random. As one symptom of leukemia is the presence of immature blood cells in the blood stream this assumption seems reasonable. The pushed-out cells are subsequently removed from the blood stream just like other cells that do not belong there. For the simulation shown in Fig 7 we have set the limit for the number of cells in the bone marrow to 30000 per microliter, roughly twice the average number of cells in the bone marrow. This number was chosen to ensure that normal steady state fluctuations in the number of cells in the bone marrow do rarely lead to cells being pushed out. The left panel of Fig 7 shows the number of normal and the number of mutated cells in the bone marrow over time. The right panels of Fig 7 show the corresponding erythrocyte, granulocyte, and thrombocyte counts in the blood. As the mutated clones completely take over the bone marrow no more normal mature blood cells and platelets can be produced and hematopoiesis breaks down completely. The break-down is very sudden due to the exponential growth of mutated cells—consistent with the sudden onset of acute leukemia. Hence, in our simulations a *single* mutation that leads to a complete differentiation stop at an early stage can lead to acute leukemia.

In order to develop acute myeloid leukemia at a later stage of the hematopoietic system a cell has to acquire more than one mutation: It has to have unlimited replicative capacity and a differentiation block. Say both mutations come together in the granulocyte lineage at stage

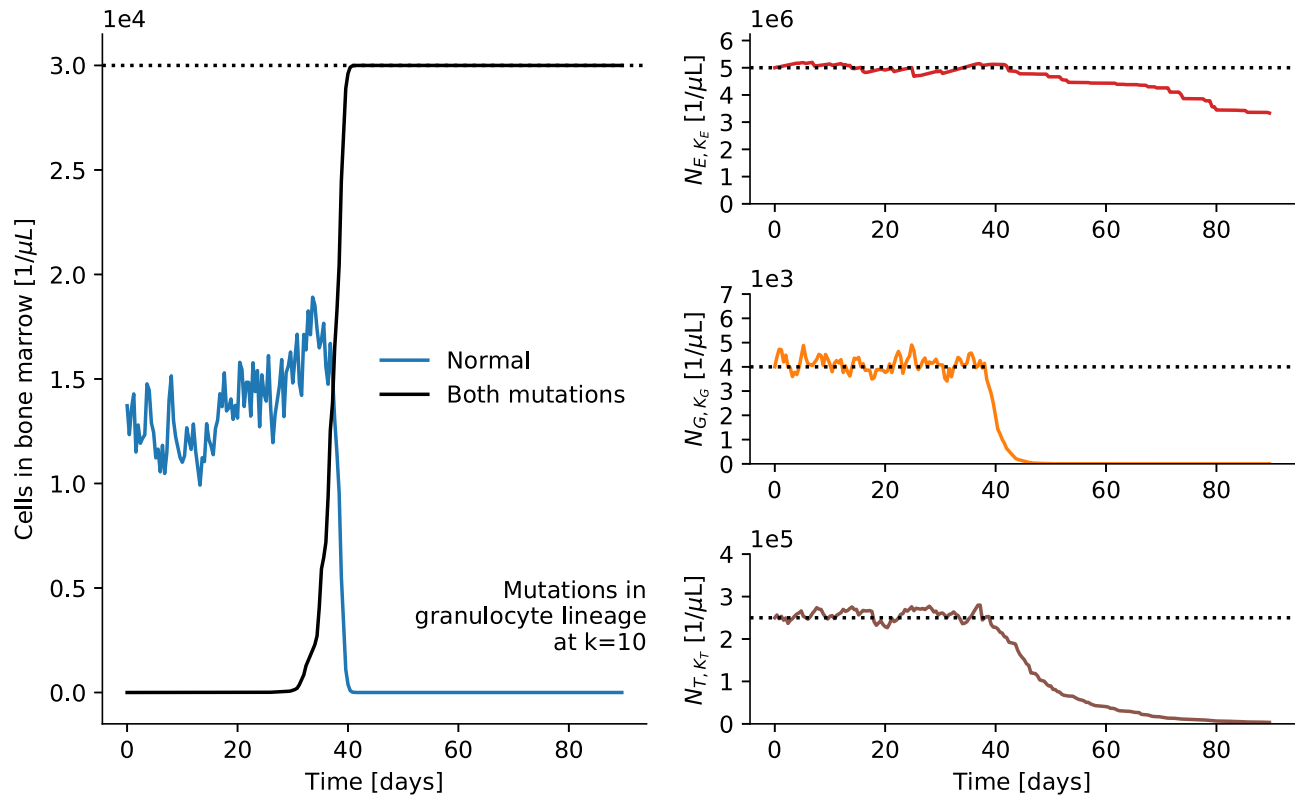


Fig 8. Late mutations. Mutations for unlimited replication and a differentiation block are introduced in the granulocyte lineage at stage $k = 10$ at time zero. The break-down of hematopoiesis happens much earlier than in Fig 7, where the mutation occurred earlier in the system.

<https://doi.org/10.1371/journal.pone.0204393.g008>

$k = 10$ and at time $t = 0$, where one mutation alone washes out quickly as shown in Fig 6). The left panel of Fig 8 shows how, after a much shorter nascent phase, the number of clones with both mutations grows extremely quickly until hematopoiesis breaks down (see right panels). Hence, if mutations for a differentiation block and unlimited replicative capacity come together, acute leukemia can also develop at later stages in our model. Still, importantly, all that is needed are two specific mutations in one cell.

The duration of the nascent phase and the growth rate of the mutated cells depend on the proliferation rate of cells at the stage where the differentiation block occurs. Hence, faster and slower growth correspond to differentiation blocks at later and earlier stages, respectively. This can also be seen when comparing the time axes of Figs 7 and 8. If limited replicative capacity is ignored, it is easy to derive an analytic expression for the expected time to diagnosis and the expected time until the complete break-down happens as a function of the stage where the mutations occur. We know the rate at which cells at that stage proliferate when the system is in steady state: $r_{x,k}$ (see panel B Fig 2). The number of cells that are added in each time step follows Eq (2) and for sufficiently small Δt the expected growth of mutated cells at this stage is therefore

$$\lim_{\Delta t \rightarrow 0} (1 + r_{x,k} \Delta t)^{\frac{t}{\Delta t}} = e^{r_{x,k} t}.$$

For a given number of mutated cells this equation can be solved for t , the time point at which this number is reached. We assume that the diagnosis is made when there are half as many mutated cells in the bone marrow as normal cells in steady state. The time until the

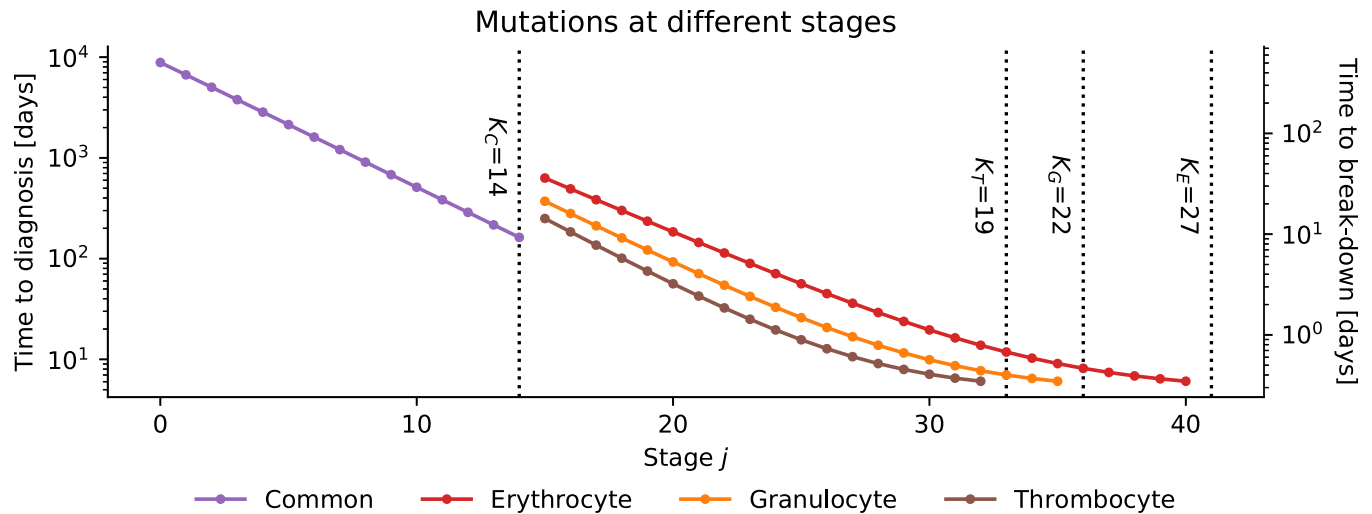


Fig 9. Time course of leukemia. Mutations for unlimited replication and a differentiation block are introduced at stage j at time zero. The time until leukemia is diagnosed is set to the time point when there are half as many mutated cells expected in the bone marrow as normal cells in steady state. This time is shown on the left y-axis. On the right y-axis the time from the diagnosis to the expected break-down of the bone-marrow is shown. This time point is assumed to be reached when there are twice as many mutated cells in the bone marrow as normal cells in steady state.

<https://doi.org/10.1371/journal.pone.0204393.g009>

diagnosis is made as a function of the stage where the mutations happen is shown on the left axis of Fig 9. Furthermore, we assume that the break-down happens when twice as many mutated cells occupy the bone marrow as normal cells in steady state. The time that passes between diagnosis and break-down is shown on the right axis of Fig 9. If at time $t = 0$ a cell with differentiation block and unlimited replicative capacity appeared at common precursor stage $k = 1$ (i.e. $j = 1$), it takes almost 20 years until the diagnosis is made. From this time point it takes about one year until the mutated cells have completely taken over the bone marrow.

One could have hoped that the only difference between acute and chronic leukemia is that chronic leukemia develops much more slowly because the mutations occur very early in the hematopoietic system. However, even for mutations that happen at the first stage of the system, one year after the diagnosis there are already twice as many mutated cells in the bone marrow than normal cells in steady state. While this is a much slower break-down of hematopoiesis than in Fig 8 it is still too fast for some cases of chronic leukemia.

Therefore, modeling chronic myeloid leukemia might require including the self-renewal mechanism of the original model of Dingli et al. [13] that we have ignored so far (cf. [16, 27–29]). The self-renewal parameter of the original model balances self-renewal and differentiation: While in our model a cell without a differentiation block always produces two daughter cells that are more differentiated than their parent, in the original model there is always the possibility that the two daughter cells are of the same type as their parent. In fact, in the original model all cells at all stages have a non-zero probability to self-renew. Hence, in the original model a differentiation block can be modeled in a graded fashion by simply increasing the probability for self-renewal at the cost of differentiation. In this way, the probability for self-renewal allows for a more fine-grained control of the growth of leukemic cells and also allows for much slower growth rates than depicted in Fig 9. Unfortunately, including self-renewal complicates the model further and requires changes to the parameter estimates, too. Furthermore, the self-renewal capabilities of non-stem-cell precursor cells are not very well understood and the assumption of the original model that all progenitors have the same capability to self-renew is at least questionable (see e.g. [30]). More experimental and theoretical work will

be required before we feel comfortable to include self-renewal in our model. Alternatively, if the number of stem cells was larger than assumed here (based on the allometric scaling arguments of [19]), the proliferation rate of the stem cells could in fact be lower, while the expected outflow of the stem cell compartment stays the same. A lower proliferation rate for the early stages could thus also make the model more consistent with the slow progression of chronic myeloid leukemia without using an additional mechanism. Another simple fix could be to assume that leukemic cells have a proliferation rate that is slower than for normal cells at the same stage.

Discussion

We have presented a model for myeloid hematopoiesis. It is based on the model of Dingli et al. [13] but extends it to explicitly include the erythrocyte, granulocyte, and thrombocyte lineages of blood formation. Considering different lineages complicates the model considerably compared to the original model. In particular, feedback in our model requires that different lineage-specific growth factor concentrations are taken into account. We have provided a conditional steady state analysis and we have shown that rough parameter estimates can be obtained from common clinical observations and readily available estimates from the literature.

For future work it will be desirable to obtain more quantitative data, i.e. beyond the blood counts for stem cell transplants already discussed, and use more principled statistical methods for parameter estimation that also produce confidence intervals (as e.g. in [17]). Such a more quantitative model should then also include the lymphoid side of blood production so that it can account for the majority of cells in a blood count.

Another direction for developing the model further is to use it to model cyclic neutropenia. The original model of Dingli et al. [13] has already been extended to capture cyclic neutropenia by also including a feedback mechanism [15], but our feedback mechanism is harder to analyze due to its non-linearity. Nevertheless, with some effort their insights can probably be transferred to our model.

Here, we have concentrated on modeling acute myeloid leukemia and extensions that seemed necessary to understand it. One such extension was the addition of a limited replicative capacity for normal cells—a property that up to now was ignored by similar models of acute myeloid leukemia (e.g. [16]). This part of the model should be made more realistic by following the lead of Werner et al. [17] who measured and modeled the distribution of telomere lengths in blood cells. They find, as expected, that with age telomere length—and therefore replicative capacity—decreases and they also show in their model that this effect can be explained by telomere shortening in the stem cell compartment. In brief, they find that telomere length decreases by roughly 50-75 base-pairs per year, which under the assumption that one cell division takes off around 50 base-pairs means that all stem cells replicate about once a year (see also [31, 32]).

A back-of-the-envelope calculation shows that the estimates that we use are, however, inconsistent with these results. We assume that there are 400 stem cells and that their proliferation rate is once a year [13, 19]. Hence, according to current estimates [17] roughly all of the 400 cells reduce their replicative potential by 1 each year. If we only had these 400 stem cells then after 30 years we could not produce any more erythrocytes as the absolute Hayflick limit is around 70 and there are about 40 stages from the stem cell compartment to the erythrocytes in our model. Luckily, the current best estimates for the overall number of hematopoietic stem cells is around 12000—of which only 400 are active at any one time (see also [19]). If we assume that the 400 active stem cells take turns, i.e. a cell that replicated is replaced by a fresh cell from the full pool, then it will take 30 years until each of the 12000 cells has divided itself

once. Hence, even after 90 years the stem cells will only have used up 3 of their 70 cell divisions given by the Hayflick limit. This is, however, inconsistent with the observed telomere shortening of around 50 base-pairs per year in mature blood cells, because this shortening implies one cell division of all cells in the full stem cell pool per year. So perhaps 400 active stem cells is too small a number and the allometric estimates on which this number is based cannot be trusted. It is thus crucial as one of the next steps to combine models for telomere shortening [17] with models of hematopoiesis and jointly estimate the number of active stem cells, their proliferation rates and the number of stages from quantitative data that include telomere distributions and blood counts. Importantly, the data that go into these models and the resulting latent parameter estimates have to be accompanied by uncertainty estimates. Otherwise it is impossible to tell whether the described inconsistencies are real or whether they are well within plausible ranges given the available data. For stochastic models like Dingli's or ours it is thus very desirable to go beyond quick-and-dirty parameter estimates and use fully Bayesian inference methods in the future (as in [17]).

Finally, as we assume in our model that leukemia is the result of mutations in specific pathways that control a cell's replicative capacity and its differentiation—in line with the conception of cancer hallmarks [8, 9, 26]—it would be desirable to link the model to genetic data. It has been suggested that only two to eight driver mutations typically lead to tumorigenesis and that only 12 signaling pathways with about 140 genes might be involved [10]. For acute myeloid leukemia these genes and mutations have recently been catalogued systematically [7] and it might hence be possible to get quantitative estimates for the probability of, for example, an acquired differentiation block.

Even without precise numbers one can make some qualitative inferences: Just like the model of Dingli et al. [13] our model suggests that many more mutations should occur at the later stages of the hematopoietic system than at the early stages, simply because there are many more cell divisions at the later stages. In fact, at about stage $j = 25$ the expected outflow is 10^7 cells per day (see panel C in Fig 2) and all these cells divide. Assuming that 10^{-7} mutations happen per gene per cell division [33] we thus expect one mutation per gene per day just for cells at this stage (cf. [34] where only stem cell divisions are considered). Luckily, mutations in just one gene at a late stage will quickly wash out as we have shown above for unlimited replicative capacity and differentiation block (see Fig 6). And mutations at an earlier stage, where the mutation has a bigger impact (see Fig 7), are less likely due to the lower number of cell division at earlier stages. Still, given that mutations should be extremely common in the hematopoietic system it seems likely that combinations of several mutations, like a differentiation block together with unlimited replicative capacity, can still occur relatively frequently at later stages. The exact probability to encounter combinations of mutations is hard to compute though as the order of acquiring the mutations matters. For example, on the one hand, as an unlimited replicative capacity washes out more quickly than a differentiation block (Fig 6) it seems more likely that the differentiation block was acquired first. On the other hand, an unlimited replicative capacity at an early stage of the system produces a huge number of descendants that can also acquire a differentiation block. Understanding such dependencies between mutations might ultimately guide the design of therapeutic protocols that target the effects of mutations in different pathways in different orders.

Acknowledgments

This research was partly funded by the Addison Fischer Institute for Research and Ethical Use of Artificial Intelligence. The authors would like to thank Theresa Göbel for combing through clinical records.

Author Contributions

Conceptualization: Sascha Lange, Roland Mertelsmann.

Formal analysis: Frank Jäkel, Oliver Worm.

Funding acquisition: Sascha Lange.

Methodology: Frank Jäkel, Oliver Worm.

Resources: Sascha Lange.

Software: Frank Jäkel, Oliver Worm.

Supervision: Sascha Lange, Roland Mertelsmann.

Writing – original draft: Frank Jäkel.

Writing – review & editing: Oliver Worm, Sascha Lange, Roland Mertelsmann.

References

1. Foley C, Mackey MC. Dynamic Hematological Disease: A Review. *Journal of Mathematical Biology*. 2009; 58:285–322. <https://doi.org/10.1007/s00285-008-0165-3> PMID: 18317766
2. Pujo-Menjouet L. Blood Cell Dynamics: Half of a Century of Modelling. *Mathematical Modelling of Natural Phenomena*. 2015; 10(6):182–205.
3. Krinner A, Roeder I, Loeffler M, Scholz M. Merging concepts—coupling an agent-based model of hematopoietic stem cells with an ODE model of granulopoiesis. *BMC Systems Biology*. 2013; 7(117): 1–20.
4. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286:531–537. <https://doi.org/10.1126/science.286.5439.531> PMID: 10521349
5. Rapin N, Bagger FO, Jendholm J, Mora-Jensen H, Krogh A, Kohlmann A, et al. Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. *Blood*. 2014; 123:894–904. <https://doi.org/10.1182/blood-2013-02-485771> PMID: 24363398
6. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *The New England Journal of Medicine*. 2015; 373:1136–1152. <https://doi.org/10.1056/NEJMra1406184> PMID: 26376137
7. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *The New England Journal of Medicine*. 2016; 374:2209–2221. <https://doi.org/10.1056/NEJMoa1516192> PMID: 27276561
8. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100:57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9) PMID: 10647931
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–674. <https://doi.org/10.1016/j.cell.2011.02.013> PMID: 21376230
10. Vogelstein B, Papadopolous N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science*. 2013; 339(6127):1546–1558. <https://doi.org/10.1126/science.1235122> PMID: 23539594
11. Stiehl T, Marciniak-Czochra A. Mathematical Modeling of Leukemogenesis and Cancer Stem Cell Dynamics. *Cancer Modeling*. 2012; 7(1):166–202.
12. Stiehl T, Baran N, Ho AD, Marciniak-Czochra A. Cell Division Patterns in Acute Myeloid Leukemia Stem-like Cells Determine Clinical Course: A Model to Predict Patient Survival. *Integrated Systems and Technologies: Mathematical Oncology*. 2015.
13. Dingli D, Traulsen A, Pacheco J. Compartmental Architecture and Dynamics of Hematopoiesis. *PLoS One*. 2007; 2(4):e345. <https://doi.org/10.1371/journal.pone.0000345> PMID: 17406669
14. Werner B, Dingli D, Lenaerts T, Pacheco JM, Traulsen A. Dynamics of mutant cells in hierarchical organized tissues. *PLoS Computational Biology*. 2011; 7:e1002290. <https://doi.org/10.1371/journal.pcbi.1002290> PMID: 22144884
15. Dingli D, Antal T, Traulsen A, Pacheco JM. Progenitor Cell Self-renewal and Cyclic Neutropenia. *Cell Proliferation*. 2009; 42(3):330–338. <https://doi.org/10.1111/j.1365-2184.2009.00598.x> PMID: 19397594

16. Werner B, Gallagher RE, Paietta EM, Litzow MR, Tallman MS, Wiernik PH, et al. Dynamics of leukemia stem-like cell extinction in acute promyelocytic leukemia. *Cancer Research*. 2014; 74:5386–5396. <https://doi.org/10.1158/0008-5472.CAN-14-1210> PMID: 25082816
17. Werner B, Beier F, Hummel S, Balabanov S, Lassay L, Orlikowsky T, et al. Reconstructing the in vivo dynamics of hematopoietic stem cells from telomere length distributions. *eLife*. 2015; 4. <https://doi.org/10.7554/eLife.08687>
18. Garcia HG, Kondev J, Orme JA N Theriot, Phillips R. A First Exposure to Statistical Mechanics for Life Scientists. arXiv.org; 2007. Available from: <https://arxiv.org/abs/0708.1899v1>.
19. Dingli D, Pacheco JM. Allometric scaling of the active hematopoietic stem cell pool across mammals. *PLoS One*. 2006; 1:e2. <https://doi.org/10.1371/journal.pone.0000002> PMID: 17183646
20. Pillay J, den Braber I, Vrsekoop N, Kwast LM, de Boer RJ, Borghans JAM, et al. In vivo labeling with ²H₂O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010; 116:625–627. <https://doi.org/10.1182/blood-2010-01-259028> PMID: 20410504
21. Harker LA, Finch CA. Thrombokinetics in man. *The Journal of Clinical Investigation*. 1969; 48:963–974. <https://doi.org/10.1172/JCI106077> PMID: 5814231
22. Catlin SN, Busque L, Gale RE, Guttrop P, Abkowitz JL. The replication rate of human hematopoietic stem cells in vivo. *Blood*. 2011; 117:4460–4466. <https://doi.org/10.1182/blood-2010-08-303537> PMID: 21343613
23. Sidorov I, Kimura M, Yashin A, Aviv A. Leukocyte telomere dynamics and human hematopoietic stem cell kinetics during somatic growth. *Experimental Hematology*. 2009; 37:514–524. <https://doi.org/10.1016/j.exphem.2008.11.009> PMID: 19216021
24. Hamey FK, Göttgens B. Demystifying blood stem cell fates. *Nature Cell Biology*. 2017; 19:261–263. <https://doi.org/10.1038/ncb3494> PMID: 28361939
25. Hjort NL. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*. 1990; 18(3):1259–1294. <https://doi.org/10.1214/aos/1176347749>
26. Groten J, Borner C, Mertelsmann R. Understanding and Controlling Cancer: The Hallmark Concept Revisited—Chance, Evolution and Entropy. *JOSHA*. 2016; 3(7):1–59.
27. Dingli D, Traulsen A, Michor F. (A)Symmetric Stem Cell Replication and Cancer. *PLOS Computational Biology*. 2007; 3(3):1–6. <https://doi.org/10.1371/journal.pcbi.0030053>
28. Traulsen A, Pacheco JM, Dingli D. Reproductive fitness advantage of BCR-ABL expressing leukemia cells. *Cancer letters*. 2010; 294:43–48. <https://doi.org/10.1016/j.canlet.2010.01.020> PMID: 20153920
29. Peixoto D, Dingli D, Pacheco JM. Modelling hematopoiesis in health and disease. *Mathematical and Computer Modelling*. 2011; 53:1546–1557. <https://doi.org/10.1016/j.mcm.2010.04.013>
30. Seita J, Weissman IL. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 2010; 2:640–653. <https://doi.org/10.1002/wsbm.86> PMID: 20890962
31. Vaziri H, Dragowska W, Allsopp RC, Thomas TE, Harley CB, Lansdorp PM. Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:9857–9860. <https://doi.org/10.1073/pnas.91.21.9857> PMID: 7937905
32. Rufer N, Brümmerdorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, et al. Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *The Journal of Experimental Medicine*. 1999; 190:157–167. <https://doi.org/10.1084/jem.190.2.157> PMID: 10432279
33. Luzzatto L, Pandolfi PP. Causality and Chance in the Development of Cancer. *The New England Journal of Medicine*. 2015; 373(1):84–88. <https://doi.org/10.1056/NEJMs1502456> PMID: 26132946
34. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. 2015; 347(6217):78–82. <https://doi.org/10.1126/science.1260825> PMID: 25554788