Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Melania Barile

born in: Isernia, Italy

Oral examination: 18 December 2018

Mathematical modelling of the kinetic properties of haematopoietic stem cells and their progeny

Contents

Summary

Haematopoiesis, the process by which blood cells are formed, is extensively studied because of its relevance for animal life. Uncovering the mechanisms of blood formation and its regulation is fundamental to cope with anomalies or illnesses such as anaemia and leukaemia, or massive blood loss.

Haematopoiesis is driven by the haematopoietic stem cells, HSCs. HSCs are able to reconstitute, upon transplantation, all blood lineages of an animal deprived of its haematopoietic cells (multipotency), and to generate one or two HSCs upon division (self-renewal). However, it is unclear how often they self-renew or differentiate into more mature compartments, according to which differentiation pathways, and how physiological and stressed conditions differ. Similarly, the kinetic properties of the progenies of the stem cells are mostly unknown.

Here we present an approach to quantify the kinetics of the haematopoietic system via a deterministic mathematical model. The model is driven by two different sets of *in vivo* measurements: fate mapping of HSCs and BrdU accumulation data.

In the first experiment we consider, an inducible, inheritable label is switched on in the stem cells without altering the physiological conditions. The fraction of labelled cells in the stem and in the downstream populations is measured over time. We build a model of population dynamics, which describes the time course increase of the labelled cells fraction in the progenies. The model has only one parameter, the time a cell resides in a population. Fitting reveals that the immediate progenies of stem cells have a long residence time, which suggest a small role of stem cells in normal haematopoiesis, sustained rather by early progenitors. We then infer the differentiation rate of a cell into its progeny by incorporating in the model the ratio of population sizes, and again confirm an infrequent contribution of stem cells.

In the second experiment we consider, the thymidine analogue BrdU is fed to mice over time. BrdU labels the cells that undergo DNA replication. The fraction of labelled cells in the stem cells and in the downstream pop-

ulations is measured over time. We adapt the population dynamics model of the previous part, incorporating the simplified assumption that cells are BrdU positive if and only if they have divided at least once. We fit the adapted model to BrdU and fate mapping data simultaneously and infer the rate at which cells divide, as well as the frequency at which division of different types (symmetric or asymmetric) happen. This analysis reveals infrequent and mainly symmetric divisions of the stem cells.

Moreover, we investigate whether a subdivision of the stem cells and their immediate progeny into several heterogeneous sub-populations is compatible with the parameters inferred as described above. We adapt the model to again fit data that consider this subdivision. We find coherent estimates for quantities that are model-invariant, which supports the robustness of our approach.

Finally, we adapt our model to describe fate mapping and cell-cycle-related data in non-stationary conditions, namely after irradiation. Contrary to normal conditions, stem cell proliferation and differentiation are significantly activated, demonstrating their importance in reconstituting a severely compromised system.

In conclusion, we suggest via data-driven deterministic modelling that HSCs fuel but do not majorly sustain normal haematopoiesis, role played by their immediate progenies. On the contrary, they are very responsive in stressed conditions, rapidly replenishing the depleted cells via enhanced proliferation and differentiation.

Zusammenfassung

Hämatopoese, der Bildungssprozess von Blutzellen, wird aufgrund seiner
Relevanz für das Leben von Organismen intensiv erforscht. Das Verständnis
der Mechanismen der Blutentwicklung und ihrer Regulation ist essentiell
um Anomalien und Krankheiten, wie Anämie und Leukämie, oder massiven
Blutverlust entsprechend behandeln zu können.

Die Hämatopoese wird von blutbildenden Stammzellen, sogenannten HSCs,
bewerkstelligt. Im Falle einer Transplanation sind HSCs dazu befähigt, alle
Blutzelltypen im depletierten Knochenmark des Empfängertieres zu regener-
ieren (Multipotenz), sowie sich selbst zu erhalten, indem sie bei der Zell-
teilung ein bis zwei identische Tochterzellen generieren (Selbsterneuerung).
Allerdings ist das genaue Verhältnis zwischen selbsterneuernden und dif-
ferenzierenden Teilungen, sowie die möglichen Entwicklungswege und beste-
hende Unterschiede zwischen physiologischen und pathogenen Bedingungen
weitgehend unklar.

Im Folgenden wird dargestellt, wie durch den Einsatz eines determinis-
tischen mathematischen Modells die dynamischen Prozesse während der
Hämatopoese quantifiziert werden können. Das Modell wird hierbei durch
die Ergebnisse zweier unterschiedlicher *in vivo* Experimente motiviert: ein-
erseits durch Messung der Zellschicksale fluoreszenzmarkierter HSCs (Fate-
Mapping), andererseits durch BrDU Akkumulationsmessungen.

Im ersten Experiment untersuchen wir einen induzierbaren, vererbbaren
Marker, der in Stammzellen angeschaltet wird, ohne physiologische Be-
dingungen zu verändern. Der Anteil an markierten Stammzellen und den
nachfolgenden Populationen wird über die Zeit gemessen. Wir erstellen
ein mathematisches Modell der Populationsdynamiken, welches den zeitlich
steigenden Anteil markierter Zellen in den Nachkommen der Stammzellen
beschreibt. Das Modell beruht auf nur einem Parameter Parameter; die
zeitliche Dauer, die eine Zelle in einer Population verbleibt (i.e. Resi-
denzzeit). Parameterschätzungen zeigen, dass die direkten Nachkommen
der Stammzellen lange Residenzzeiten haben. Dies suggeriert eine unterge-
ordnete Rolle der Stammzellen bei normaler Hämatopoese, die stattdessen

aufrechterhalten wird von ersten Nachkommen. Wir sagen dann Differenzierungsraten von Zellen zu deren Nachkommen vorher, indem wir die Verhältnisse der Populationsgrößen einbeziehen, was wiederum die seltene Beteiligung der Stammzellen bestätigt.

Im zweiten Experiment werden Mäuse mit dem Thymidin-Analogon BrdU gefüttert, das proliferierende Zellen markiert. Der Anteil der markierten Zellen innerhalb der Stammzellpopulation und der Nachfolgepopulationen wird über die Zeit hinweg gemessen. Wir verändern das zuvor generierte Modell der Populationsdynamik, indem wir die vereinfachte Annahme einbauen, dass Zellen nur dann BrdU-positiv sind, wenn sie sich mindestens einmal geteilt haben. Um die Häufigkeit der (symmentrischen oder assymetrischen) Zellteilungen zu bestimmen, passen wir das Modell an die BrdU- und Fate-Mapping-Daten an. Die Analyse zeigt seltene und hauptsächlich symmetrische Teilungen der Stammzellen.

Des Weiteren untersuchen wir, ob eine Unterteilung der Stammzellen und ihrer unmittelbaren Nachkommen in verschiedene heterogene Subpopulationen mit den gelernten Paramtern vereinbar ist. Daraufhin adaptieren wir das Modell erneut, um es den Daten anzupassen, die diese Unterteilung betreffen. Wir finden dabei kohärente Schätzungen für modell-unabhängige Mengen, was wiederum die Robustheit unseres Ansatzes bestätigt.

Schließlich passen wir unser Modell so an, dass wir Fate-Mapping und Daten des Zellzyklus unter nicht natürlichen Bedingungen (besonders nach Bestrahlung) beschreiben können. Im Vergleich zu natürlichen Bedingungen werden Proliferation und Differenzierung signifikant erhöht. Damit wird ihre Bedeutung für die Wiederherstellung eines schwer beeinträchtigten hämatopoetischen Systems untermauert.

Zusammenfassend: Durch datengetriebene, deterministische Modellbildung gelangen wir zu dem Schluss, dass HSCs die normale Hämatopoese versorgen, aber nicht hautsächlich aufrechterhalten; diese Rolle fällt ihren unmittelbaren Nachfolgerpopulationen zu. Dennoch reagieren HSCs empfindlich auf Stressbedingungen, indem sie verlorene Zellen durch erhöhte Proliferation und Differenzierung schnell ersetzen.

Chapter 1

Introduction

Haematopoiesis

The formation of blood cells is named "haematopoiesis" after the greek *ema* (blood) and *poiesis* (formation). Blood cells are formed throughout all the life time of an animal. In mice, the first haematopoietic cells appear in the yolk sac at day 7 of gestation. The site of haematopoiesis then migrates to the aorta-gonad mesonephros, the placenta, and from embryonic day 10 to the foetal liver. After birth haematopoietic stem cells colonise the bone marrow and drive adult haematopoiesis [Golub and Cumano, 2013]. In adult vertebrates, thereafter, the system maintains the cellularity of blood and immune cells over the life span of the animal. The cells that are most rapidly replenished are neutrophils, platelets and erythrocytes, whereas lymphocytes have longer lifetimes. Cell replenishment adapts to heightened demand caused, e.g., by infection or blood loss.

Haematopoietic stem cells

A central role in haematopoiesis is played by haematopoietic stem cells (HSCs). The definition of HSCs dates back to the 1950s, and it has been developed on the basis of experiments performed on animals exposed to radioactive irradiations. Since irradiation kills stem and progenitor cells (that is, dividing cells), animals would not survive the exposure to high doses. On the other hand, animals receiving haematopoietic progenitors via transplantation did survive, and the composition of their blood was fully restored, meaning that mature blood cells were produced by the transplanted cells *de novo* ([Ford et al., 1956], [Nowell et al., 1956]).

Over the years, the phenotype of the pool of cells that must be transplanted to guarantee the recovery of the animals was more and more specifically determined. Such cells were named stem cells and it was shown that even

a single stem cell can repopulate the whole system [Ema et al., 2005]. Furthermore, it was checked whether such reconstitution was stable over time. Transplantation of some progenitor did not guarantee a durable survival, while upon transplanting stem cells the number of blood cells did not decline over time.

Also, it was noticed that the daughter cells generated by a symmetric division of a transplanted stem cells could reconstitute an irradiated animal upon secondary transplantation, meaning that the stem cells can produce cells that have the same multipotent reconstituting potential. In literature, such capability is referred to as self-renewal. Evidence for self-renewal of HSCs has been provided by retroviral marking studies in which HSC clones tagged with retroviral integration sites were transplanted into secondary recipients ([Dick et al., 1985], [Keller et al., 1985], [Keller and Snodgrass, 1990], [Lemischka et al., 1986]).

In conclusion, it has been known for decades that there are some special cells that are able to self-renew and differentiate to produce all the mature blood cells, but many open questions about the characterisation of such cells and their progeny, and the architecture and regulation of the haematopoietic tree, still persist.

Current knowledge and open questions

Characterisation of the stem cells

A major point of debate is the phenotypical characterisation of haematopoietic stem cells. The current standard is to isolate HSCs based on the surface markers Lin^- Kit^+ Sca^+ $CD150^+$ $CD48^-$ [Kiel et al., 1995], but within this basic characterisation the existence of distinct sub-compartments has been proposed by different groups.

For example, in [Oguro et al., 2013], HSCs are further subdivided according to the additional surface marker CD229, with the $CD229^-$ population being the immature HSC-1, which give rise to the less immature $CD229^+$ population, HSC-2. This hierarchy was defined based on reconstitution and proliferative potential.

In a more recent publication , the HSCs are purified based only on the expression of the homeobox B5 (Hoxb5) gene, and $Hoxb5^+$ cells only are able to provide long term reconstitution [Chen et al., 2016]. On the other end, [Ito and Ito, 2016] suggest that *Tie2* expression discriminates among HSCs, with positive cells being the most immature stem cells.

Other studies indirectly identify a most primitive population within the HSCs, although they provide neither a phenotypical nor a molecular characterisation. In [Sawai et al., 2016] authors performed a fate mapping of HSCs and found that they must be heterogeneous since, if labelled with a bacterial artificial chromosome (BAC) clones driving transgene expression from the Pdzk1ip1 locus, their labelling frequency increases over time. In [Wilson et al., 2008] cell cycle label retention suggests HSCs to be heterogeneous with respect to their proliferative potential.

[Benz et al., 2012] and [Luchsinger et al., 2016] suggest that HSCs are heterogeneous based on clonal analysis post-transplantation, which yielded that cells are biased towards the production of different progenies. However, transplantation assays are a potentially different situation from physiological haematopoiesis.

Topology

Much work has been dedicated to assessing the differentiation hierarchy, or topology, from stem cells to mature populations.

In the classical view of haematopoiesis, HSCs are long term reconstituting and progressively lose capability of self-renewal upon differentiating into short term reconstituting haematopoietic stem cells (ST-HSCs) and then further into multipotent progenitors (MPPs) [Morrison and Weissman, 1994]. MPPs then commit to be either myeloid cells upon differentiating into common myeloid progenitors, CMPs [Akashi et al., 2000], or to be lymphoid cells turning into common lymphoid progenitors CLPs [Kondo et al., 1997]. CMPs and CLPs subsequently lose multipotency in a stepwise manner to form all mature cell types. This model has been based on transplantation studies and colony assays. However, a recent paper using barcodes and fate mapping of endogenous HSCs supports a principal dichotomy between myelo-erythroid and lymphoid developments [Pei et al., 2017].

Several works, however, question this model and postulate the existence of alternative or supplemental routes [Perié et al., 2014]. To form erythrocytes, for example, direct pathways might emerge from CMPs [Nutt et al., 2005], MPPs [Lai and Kondo, 2006] or earlier from stem cells [Guo et al., 2013] [Notta et al., 2016] [Rodriguez-Fraticelli et al., 2018], although in [Boyer et al., 2010] authors proved the necessity of an intermediate FLK2+ progenitor between HSCs and all mature cells. [Hoppe et al., 2016] question the early myeloid choice, and [Yamamoto et al., 2013] and [Perié et al., 2015] that the MPP population is required in the developmental pathway at all.

Note that topology is strictly connected with the pre-discussed heterogene-

ity issue. In fact, once there is evidence that a population at any point of the differentiation chain is heterogeneous, then the question about the existence of multiple differentiation pathways also arises. Particularly, several publications claim that population ST-HSC and/or MPP are heterogeneous [Oguro et al., 2013], and propose that each subpopulations is biased toward the fate of a specific mature blood cell via a lineage tracking experiment [Pietras et al., 2015].

Quantitation of haematopoiesis

Moreover, the scientific community has been interested in estimating kinetic parameters such as the frequency at which HSCs differentiate, proliferate or die in normal conditions[1].

Estimating the proliferation rate is easier because one only needs to collect cell-cycle-related data on HSCs. In several works, for example, the thymidine analogue BrdU is administrated to the animals via drinking water. BrdU is incorporated in the DNA during the S phase of the cell cycle and thus labels cycling cells. The fraction of stem cells becoming BrdU positive is measured over time as in [Oguro et al., 2013], and the fraction of recruitment into cell cycle estimated via modelling (for example, once in 12 days as estimated in [Cheshier et al., 1999]).

[Mackey, 2001] proposed a model that describes the dynamics of BrdU accumulation over time. This model assumes a fixed duration of the cell cycle, but a stochastic exit from the G_0 phase, either by commitment into cycle or by cellular differentiation. Calibrating the model on BrdU continuous labelling data measured by [Bradford, 1997] and [Cheshier et al., 1999] yields that proliferation occurs once in 20 to 50 days. Another consequence of this model is that the number of proliferations a HSC undergoes before fully maturing is on average 20. Interestingly, the same model provides also an upper bound for differentiation and death rate (once in 50 or 2 days respectively). As we will show in this thesis, it is only possible to obtain also a lower bound on the estimated differentiation and death rates of a population if information on the kinetics of its direct progeny is also available.

It has been suggested that BrdU can have a toxic effect, and thus the potential to activate stem cells. To circumvent this problem, measurement of the fraction of BrdU can be performed in the delabelling phase following BrdU removal as in [Wilson et al., 2008], [van der Wath RC, 2009]. Note that this strategy assumes that shortly after BrdU removal the physiological conditions are restored. Chase experiments were also performed by [Foudi

---

[1]Here we exclusively report results obtained with mouse models. Kinetics were proven to be different depending on the animal size [Abkowitz et al., 2000]

et al., 2008] regaridng the dilution of the histone 2B-GFP protein. Altogether, these works found, via deterministic or stochastic modelling, that stem cells divide on average once in 100 days.

In [Abkowitz et al., 2000], authors performed stochastic modelling of the clonal dynamics of terminally differentiated populations during repopulation after transplantation. They concluded that, in the stem cell population, a proliferation and a differentiation event take place on average once every 20 days, and that a death event occurs way less frequently. An infrequent activity by stem cells and a major role of their progeny in sustaining haematopoiesis is also the conclusion of [Schoedel et al., 2016], who ablated HSCs and saw no major effect on their progeny for at least one year. On the other hand, in [Sawai et al., 2016] authors have estimated a more frequent commitment (ones in 10 days) for the stem cells based on mathematical modelling of HSCs fate mapping data.

Finally, few data exist on the kinetic parameters of populations other than stem cells, apart for the commitment rate of progenitors of T and B cells [Tough and Sprent, 1994] [Rauch et al., 2009] and the death rate of a few terminally differentiated populations [Vácha et al., 1980] [van Furth R, 1989]. Some of these gaps have been filled by theoretical inference upon simulating the behaviour of the haematopoietic system in stationary situations and after challenge [Manesso et al., 2013] [Marciniak-Czochra et al., 2008].

Mechanisms of proliferation

Another point of interest is the mechanism of proliferation and differentiation of stem cells. Experimental evidence exists that cells can divide symmetrically or asymmetrically [Wu et al., 2007], but the estimation of the relative probability of such events is contradictory [Yamamoto et al., 2013], [Ito and Ito, 2016], [Lai and Kondo, 2006] [Shin et al., 2014], [Morrison and Kimble, 2006], [Marciniak-Czochra et al., 2008].

Number of stem cells contributing to haematopoiesis

Finally, many studies have addressed the quantification of the number of HSCs that contribute to haematopoiesis via clonal analysis of individually marked HSCs transplanted in irradiated mice. These experiments yielded that only a small fraction of HSCs contribute to haematopoiesis, in the order of tenths of percents [Naik et al., 2013] [Gerrits et al., 2010] [Lu et al., 2011] [Jordan and Lemischka, 1990]. On the other hand, it has been shown that, in normal conditions, many clones differentiate into committed progenitors [Sun et al., 2014], and that nearly all HSCs proliferate within the life span of a mouse, as shown by label retaining assays in [Bernitz et al., 2016] or

[Cheshier et al., 1999].

In this thesis

A simple model for normal haematopoiesis

In this chapter we build a mathematical model to quantify the differentiation kinetics and the time a cell spends in a population (residence time) based on fate mapping data. The fate mapping experiment was performed by Dr. Katrin Busch [Busch et al., 2015]. Dr Busch developed a knock-in mouse that allows to induce a label in the solely HSCs in a permanent, genetically transmissible fashion (Figure 1.1). When the mice are around 6 weeks old,
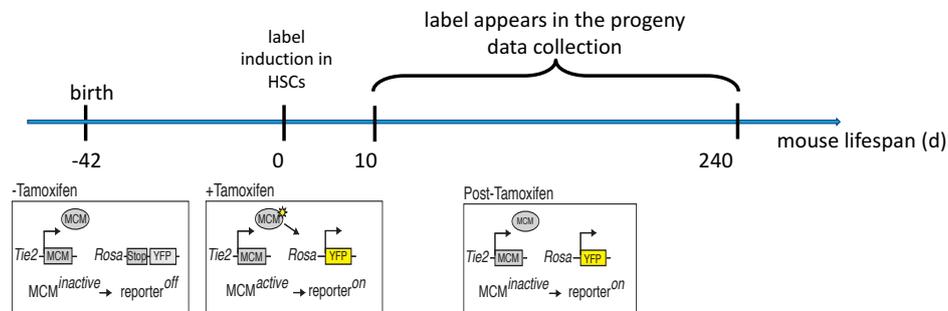
Figure 1.1: Experimental scheme of HSCs' fate mapping [Busch et al., 2015]. Mutant mice expressing MCM (an improved Cre fused with modified oestrogen receptor) from the *Tie2* locus are crossed with reporter mice with a YFP protein in the *Rosa* locus. Left panel: in the absence of tamoxifen, MCM is inactive and the YFP is not expressed. Middle panel: when the mice are 6 weeks old, tamoxifen is administrated, and MCM activated as a consequence. MCM then migrates to the nucleus and recombines the *Rosa* locus, removing the Stop codon. From this moment on, the YFP is expressed (right panel).

a genetic yellow fluorescent label is permanently switched on in the HSCs exclusively. Around five week afterwards the first labelled progenies appear, and their number increases over time. In this first part, the standard definition of populations and their homogeneity are not questioned. From this approach we learn differentiation, net proliferation rates and residence times of several haematopoietic populations. We additionally estimate the minimum number of HSCs that actively contribute to haematopoiesis during the experimental time span and we gain some knowledge about the developmental topology.

Refining the model: heterogeneity, proliferation

In this chapter we aim at disentangling the proliferation and death rates by means of additional data and an appropriate mathematical modelling approach. The data are derived from a BrdU assay performed by [Oguro et al., 2013]. When the mice are around 8 weeks old, BrdU is administrated and all
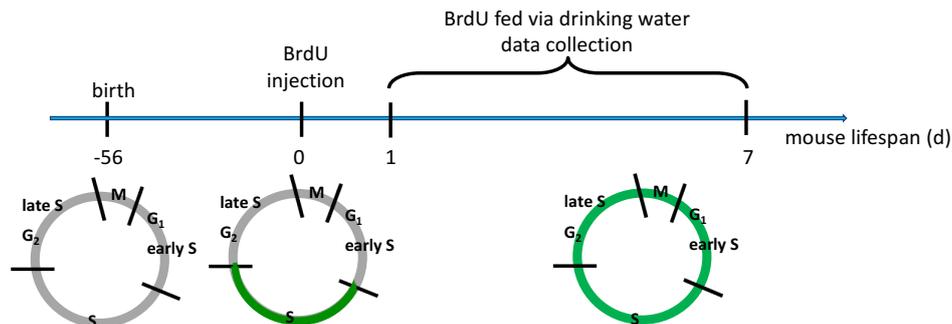
Figure 1.2: Experimental scheme of BrdU labelling of murine cells [Oguro et al., 2013]. Segments in a circle represent the fraction of cells in the corresponding cycle phase. Gray=non labelled, green=labelled.

the cells that are replicating DNA (S phase) become labelled. Since BrdU is continuously fed to the mice, progressively all cells that go through S phase become labelled, without label dilution. The model describes the kinetics of BrdU accumulation over time. In this chapter we also investigate the heterogeneity of stem cells populations and the mechanisms of proliferation that are inferable from the very same label propagation data.

Comparing estimated parameters among different models

In this chapter we present a theoretical framework to compare different models that consider the same cellular populations homogeneous and heterogeneous respectively. We show that the values of parameters such as differentiation or proliferation depend on the model, but there exist quantities that do not, such as resident time, number of proliferations before leaving a compartment, flux into a compartment.

Non-stationary haematopoiesis after irradiation

Finally, we build a model to describe non steady state haematopoiesis, particularly after sub-lethal irradiation. Ann-Katrin Schuon measured the total number of cells and the labelling frequency over time up to 16 weeks after irradiation for the HSC compartments and its progeny (unpublished data), plus the EdU accumulation and dilution at different time points. We
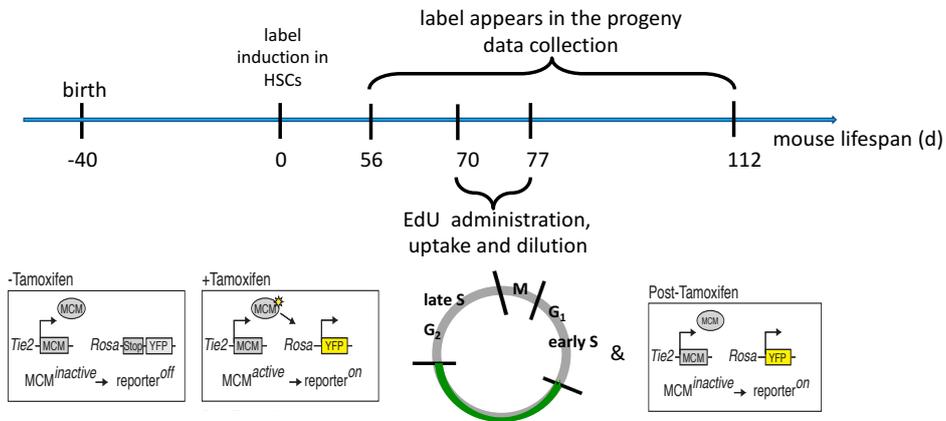
Figure 1.3: Experimental scheme of fate mapping combined with EdU labelling of murine cells. Experiment performed by Ann-Kathrin Schuon (unpublished). Segments in a circle represent the fraction of cells in the corresponding cycle phase. Gray=non labelled, green=labelled

compute the cell cycle length, differentiation and death rates for stem and progenitor cells and we compare it to stationary case.

Chapter 2

A simple model for normal haematopoiesis

2.1 HSC activity

The first point we addressed was the quantification of the contribution of the HSCs and immediate progenitors to normal haematopoiesis. More precisely, we posed the following questions:

1. In case of HSCs failure, would we see visible effects on the progeny in the near term?

2. How often does an HSC produce a direct progeny?

3. How many HSCs contribute to haematopoiesis?

These questions require a quantitative approach, that is, a mathematical model, to be tailored to suitable data.

2.1.1 The experimental framework

We use the data generated by our experimental partners [Busch et al., 2015] via fate mapping of the HSCs. The idea was to generate a mouse model where a genetic and hence permanent label is exclusively induced in the HSCs and to measure, over time, the fraction of labelled cells in all populations (also called compartments in this work). The requirement of a genetic (and thus inheritable) label is fundamental for the questions we address, since they all involve characterisation of HSCs via output on their progeny. A permanent label is important to observe the system over long times without having to introduce a model for the dilution of the label. Finally, exclusively labelling HSCs was fundamental to address point 3, as will be explained in Section 2.1.5.

The labelling was performed in a knock-in mutant expressing, from the *Tie2* locus, a gene encoding codon-improved Cre (iCre) fused to two modified oestrogen receptor binding domains (designated as MCM). The $Tie2^{MCM}$ allele was crossed to $Rosa^{YFP}$ mice, expressing the yellow fluorescent protein (YFP) reporter in a Cre-dependent manner. After tamoxifen treatment, MCM becomes active and deletes the stop cassette of the YFP marker gene, thus rendering Cre-expressing cells and their non-Cre-expressing progeny YFP-positive. See Figure 2.1.1 for the scheme of the experiment. Due to
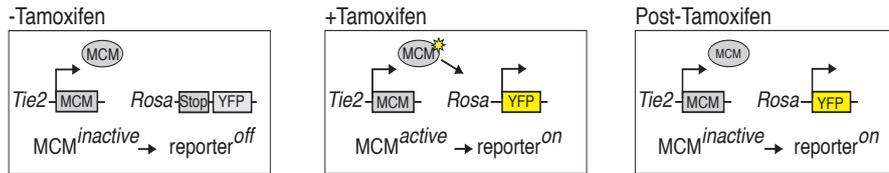
Figure 2.1: Switching on a heritable label in the HSCs. Left: In the absence of tamoxifen there is no YFP expression. Middle: If tamoxifen is added, the MCM is activated, and recombines the *Rosa* locus, removing the stopping codon and allowing the expression of the YFP. Right: Even long after tamoxifen administration and though MCM is returned to the inactive state, YFP remains expressed [Busch et al., 2015].

the specificity of *Tie2* expression, the described procedure exclusively tags HSCs. There is a very small proportion (less than 0.01%) of labelled cells in the myeloid compartments, but their life time is so short that this tiny fraction is lost rapidly. In the following weeks, the labelled HSCs undergoing differentiation start invading the downstream compartments, thus increasing the labelling frequency in the progeny (Figure 2.2). The frequency of labelled cells over time for all the populations of interest is the observable we are interested in modelling.

2.1.2   Model basis

In order to be able to specifically address the above-mentioned quantitative questions we now define the cell-fate-related parameters and the basic ideas behind modelling population kinetics. In Figure 2.3 we present all the possible fates a cell can undergo. A is a general progenitor compartment and B is its direct progeny. Rates below each scheme denote the number of events of that kind between t and $t + dt$ per dt per cell of type A. Given such definitions, the time evolution of the number of cells $T_A(t)$ and $T_B(t)$
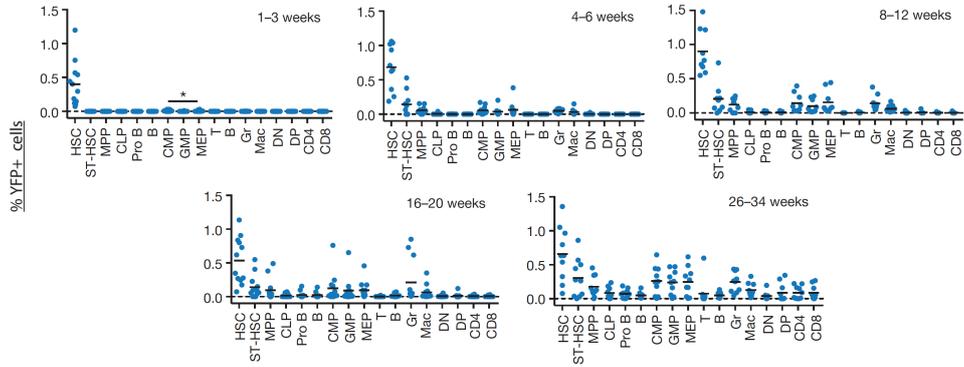
Figure 2.2: Label propagation as measured by [Busch et al., 2015]. The average percentage of YFP$^+$ cells in each measured compartment is plotted over time (see Session Abbreviation). Blue dots represent individual mice (n=114).
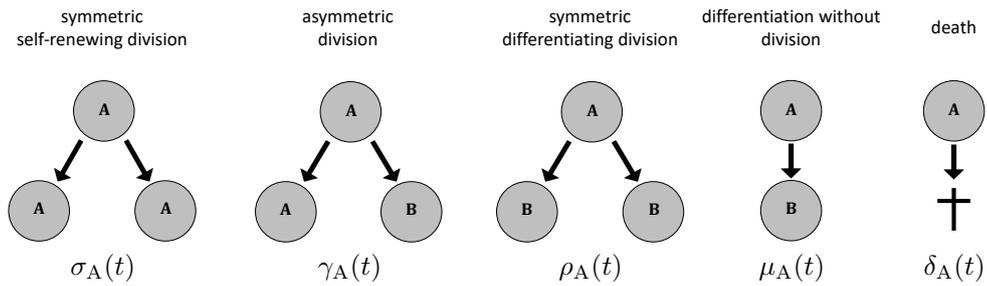
Figure 2.3: If A is a generic cell population upstream of B, then there are 5 possible fates A can undergo with rates $\sigma_A(t)$, $\gamma_A(t)$, $\rho_A(t)$, $\mu_A(t)$ and $\delta_A(t)$ respectively.

in populations A and B, respectively, is [Flossdorf, 2013]:

$$\dot{T}_A(t) = (\sigma_A(t) - \rho_A(t) - \mu_A(t) - \delta_A(t))\, T_A(t) \tag{2.1}$$

$$\dot{T}_B(t) = (2\,\rho_A(t) + \gamma_A(t) + \mu_A(t))\, T_A(t) + \tag{2.2}$$

$$+ (\sigma_B(t) - \rho_B(t) - \mu_B(t) - \delta_B(t))\, T_B(t) \tag{2.3}$$

We assume that, under the steady state assumption, all rates are constant. It is straightforward to see that for this kind of problem the parameters used are redundant. Upon substituting:

$$\lambda_A = \sigma_A + \rho_A + \gamma_A \tag{2.4}$$

$$\alpha_A = 2\,\rho_A + \gamma_A + \mu_A \tag{2.5}$$

we obtain the equivalent system:

$$\dot{T}_A(t) = (\lambda_A - \alpha_A - \delta_A)\, T_A(t) \tag{2.6}$$

$$\dot{T}_B(t) = \alpha_A\, T_A + (\lambda_B - \alpha_B - \delta_B)\, T_B(t) \tag{2.7}$$

We can summarise the meaning of these substitutions in the following definitions:

- $\lambda_A(t)$ is the proliferation rate of population A. This is the number of proliferation events among the cells of type A between t and $t + dt$ per cell of type A per dt:

$$\lambda_A(t) = \frac{\#\text{cells of type A that proliferate between } t \text{ and } t + dt}{dt\, T_A}$$
$$\tag{2.8}$$

- $\alpha_A(t)$ is the differentiation rate of population B. This is the number of cells of type B that are formed by A between t and $t + dt$ per cell of type A per dt:

$$\alpha_A(t) = \frac{\#\text{cells of type B that are formed by A between } t \text{ and } t + dt}{dt\, T_A}$$
$$\tag{2.9}$$

- $\delta_A(t)$ is the death rate of population A. This is the number of cells of type A that die between t and $t + dt$ per cell of type A per dt:

$$\delta_A(t) = \frac{\#\text{cells of type A that die between } t \text{ and } t + dt}{dt\, T_A} \tag{2.10}$$

Note that the two parameterisations are equivalent only when it comes to describe the time evolution (and thus the steady-state ratio) of the number of cells in different populations. We will see in Section 3 that this is not true in general.

We are now ready to describe the modelling approach for each of the posed questions.

2.1.3 Effects of HSCs failure

The first question we posed is:

"In case of HSCs failure, would we see visible effects on the progeny in the near term?"

In order to achieve a quantitative answer, we can think of the following experiment: Let us imagine that we are able to suddenly ablate all HSCs. Furthermore, we assume that this process does not change the kinetic rates of ST-HSCs, that is, no feedback mechanism would be triggered by the ablation. ST-HSCs would start declining immediately after the input is cut out, and so will all downstream populations. We aim at quantifying the speed of the decline. This is equivalent to pose the question:

"How long does it take, on average, to a cell of type ST-HSC to exit the compartment because of either differentiation or death?"

Thus, we are aiming at estimating a quantity, say τ_{ST}, called from now on the residence time, which represents the average time spent by a cell and its progeny in the ST-HSC compartment. A short residence time means that the effect of HSCs ablation would be seen immediately in the haematopoietic output, thus implying a major relevance of HSC contribution to normal haematopoiesis.

Note that for the question to be correctly posed the rate at which cells are lost from the ST-HSC compartment must exceed the rate at which cells are produced, otherwise we would be dealing with an expanding population, for which the residence time would be infinite.

The rate of variation over time of the number $T_{ST}(t)$ of cells of type ST-HSC at time t is defined, as in equation 2.7, as:

$$\dot{T}_{ST}(t) = \alpha_{HSC}\, T_{HSC}(t) + (\lambda_{ST} - \alpha_{ST} - \delta_{ST})\, T_{ST}(t) \qquad (2.11)$$

Now, if the input is ablated, we obtain:

$$\dot{T}_{ST}(t) = (\lambda_{ST} - \alpha_{ST} - \delta_{ST})\, T_{ST}(t) = -\kappa_{ST}\, T_{ST}(t) \qquad (2.12)$$

where $\kappa_{ST} = \alpha_{ST} + \delta_{ST} - \lambda_{ST}$ is the net efflux from ST and thus

$$T_{ST}(t) = e^{-\kappa_{ST}\, t} \qquad (2.13)$$

The average time τ_{ST} for a cell to be lost is by definition:

$$\tau_{ST} = norm \int_0^\infty t\, (\alpha_{ST} + \delta_{ST})\, T_{ST}(t)\, dt \qquad (2.14)$$

where $(\alpha_{\mathrm{ST}} + \delta_{\mathrm{ST}})\, T_{\mathrm{ST}}(t)\, dt$ is the probability of having either a differentiation or death event between t and $t + dt$ and the normalisation factor $norm$ is:

$$norm = \int_0^\infty (\alpha_{\mathrm{ST}} + \delta_{\mathrm{ST}})\, T_{\mathrm{ST}}(t)\, dt \tag{2.15}$$

Equations 2.13, 2.14 and 2.15 yield $\tau_{\mathrm{ST}} = \dfrac{1}{\kappa_{\mathrm{ST}}}$ and thus we need to estimate κ_{ST}. As we anticipated, $\kappa_{\mathrm{ST}} > 0$ because cell loss has to exceed cell production.

The observables of the fate mapping experiment described in Section 2.1 are, for each mouse and each cell type of interest, the total number of cells in a certain aliquot and the number of labelled cells in the same aliquot. The total number of cells in a population is assumed to be constant at steady state (Figure 8.1 of the Supplementary Material). while the number of labelled cells grows over time (Figure 2.2). labelled cells and unlabelled cells are not treated differently from the point of view of the rates. For the HSC compartment, we assume perfect self-renewal, that is, the proliferation rate perfectly balances the death and the differentiation rates and this means that not only the total number of HSCs is constant over time but also that of labelled HSCs. The perfect self-renewal assumption holds because the HSCs are at the top of the differentiation hierarchy, so they do not have any input and need to sustain the flux into haematopoiesis. They could in principle grow over time, as considered in the Supplementary Material (Figure 8.2)

Upon denoting $L_{compartment}(\mathrm{t})$ the time function of the number of labelled cells in a compartment, we can write:

$$\dot{L}_{\mathrm{ST}}(t) = \alpha_0\, L_{\mathrm{HSC}} - \kappa_{\mathrm{ST}}\, L_{\mathrm{ST}}(t) \tag{2.16}$$

Due to experimental design, the only quantity we can compare among different animals is the fraction of labelled cells in an aliquot, $r_{compartment}(t)$. From

$$\dot{r}_{\mathrm{ST}}(t) = \frac{\dot{L}_{\mathrm{ST}}(t)}{T_{\mathrm{ST}}} = \frac{\alpha_{HSC}\, L_{\mathrm{HSC}}(t)}{T_{\mathrm{ST}}} - \kappa_{\mathrm{ST}}\, r_{\mathrm{ST}}(t) \tag{2.17}$$

Now, for the total numbers of cells it is true that:

$$\dot{T}_{\mathrm{ST}} = \alpha_{\mathrm{HSC}}\, T_{\mathrm{HSC}} - \kappa_{\mathrm{ST}}\, T_{\mathrm{ST}} = 0 \tag{2.18}$$

Upon combining 2.17 and 2.18 we find:

$$\dot{r}_{\mathrm{ST}}(t) = \kappa_{\mathrm{ST}}\, (r_{\mathrm{HSC}} - r_{\mathrm{ST}}(t)) \tag{2.19}$$

In other words, the time course of the fraction of labelled cells permits to directly estimate the residence time. Note that the last equation also states

14

that the labelling frequency in ST-HSC tends to equilibrate with the one in HSC, and thus, since the initial condition is 0, it asymptotically grows. Equation 2.19 can be further simplified upon dividing left and right hand sides by the constant r_{HSC}[1]. This allows to elegantly write the time evolution of the fraction of labelled cell in population ST normalised by the labelling frequency in HSC as:

$$\dot{f}_{\mathrm{ST}}(t) = \kappa_{\mathrm{ST}} \left(1 - f_{\mathrm{ST}}(t)\right) \qquad (2.21)$$

This simplification is not necessary for modelling purposes, but it allows to reduce the parameter numbers (no need to estimate the labelling frequency of HSC) and thus to focus on modelling ST-HSC only. We indeed reached similar conclusions upon modelling the normalised labelling frequency of ST-HSCs or the labelling frequency of HSC and ST-HSC. Further on we will question the homogeneity of the HSC compartment and will need to fit their initial labelling frequency as well (see Section 3.1).

The model as described above successfully fits the data (details in Methods):
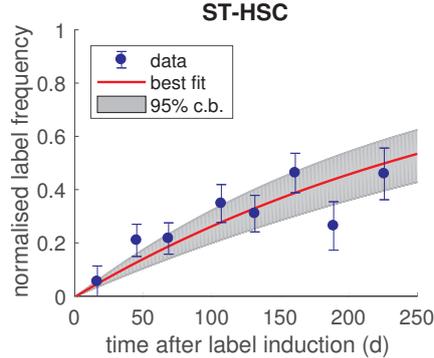
Figure 2.4: Model reproduces the time increase of the labelling frequency in population ST-HSC. Grey shade: 95% confidence bounds as calculated via bootstrap simulations (see Methods).

We found that:

$$\kappa_{\mathrm{ST}} = 0.003 \ [0.002 - 0.004] d^{-1}$$

---

[1]We saw that, in steady state, perfect self-renewal of the HSC compartment implies that both L_{HSC} and T_{HSC} are constant over time, and thus so is their ratio. But a constant labelling frequency is actually true for a population X without any input (like the case of the most immature stem cells) even in non-stationary conditions and whatever the time dependance of the rates is:

$$\dot{r}_{\mathrm{X}}(t) = \frac{1}{T_{\mathrm{X}}}(\lambda_{\mathrm{X}}(t) - \alpha_{\mathrm{X}}(t) - \delta_{\mathrm{X}}(t)) \, L_{\mathrm{X}} - \frac{L_{\mathrm{X}}}{T_{\mathrm{X}}^2}(\lambda_{\mathrm{X}}(t) - \alpha_{\mathrm{X}}(t) - \delta_{\mathrm{X}}(t)) \, T_{\mathrm{X}} = 0 \qquad (2.20)$$

or, analogously,
$$\tau_{\mathrm{ST}} = 330 \ [250 - 500]d$$
where the values in brackets represents the 95% confidence intervals on the best fit estimation as calculated with the profile likelihood method (see Methods and Figure 8.3 of the Supplementary Material).

We have now an answer for our initial question: a direct progeny of an HSC, a cell of type ST-HSC, resides in its compartment for around one year. This means that, for around one year, the effects of a sudden loss of HSCs are not measurable on the mature haematopoietic compartments, which all come from the ST-HSCs, and thus the contribution of the HSCs to normal haematopoiesis is not detectable for time scales of around one year (note: this conclusion only holds if the HSCs exclusively feed the ST-HSC compartment; more details about alternative lineage topologies are given in 2.3).

A simulated effect of HSC ablation is plotted in Figure 2.5.

2.1.4 Frequency of differentiation

We assessed in the previous Section the key role of the ST-HSCs in sustaining haematopoiesis. It could still be, though, that the ST-HSCs themselves needed a massive input from HSCs to guarantee their compartment size.

From Equation 2.18, we have:

$$\alpha_{\mathrm{HSC}} = \kappa_{\mathrm{ST}} \frac{T_{\mathrm{ST}}}{T_{\mathrm{HSC}}} \tag{2.22}$$

Since we have already estimated the efflux rate, we only need to measure the ratio of compartment sizes $\frac{T_{\mathrm{ST}}}{T_{\mathrm{HSC}}}$ to estimate α_{HSC}[2].
We found that:
$$\alpha_{\mathrm{HSC}} = 0.01 \ [0.007 - 0.013]d^{-1}$$

Note: again, this result for HSC differentiation is only true if the HSCs exclusively feed the ST-HSC compartment. In fact, it might in principle be that HSC differentiate into a certain compartment X. The frequency of this hypothetical differentiation event is, with the data described so far, not distinguishable from the death rate, for which we do not have so far a direct estimate. In fact, having estimated the flux into one progeny yields

---

[2]In reality, the adopted procedure is to fit all the data together, label propagation and ratio of compartment sizes. This is more rigorous, since the parameters are estimated in such a way that they reproduce all data simultaneously.
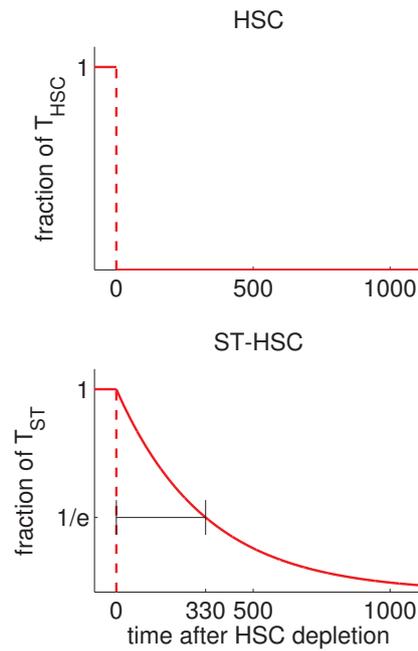
Figure 2.5: Effects of HSCs ablation. Before time 0 days both populations HSC and ST-HSC are at steady state (normalised value 1). After HSCs ablation ($t > 0$), the number of ST-HSCs declines slowly, but for around one year its size is still circa 40% of the stationary value, implying that HSCs ablation does not massively affect haematopoiesis for a long time span. This simulation assumes that HSCs ablation leaves the residence time of ST-HSCs unaffected.

an estimate for the difference between proliferation rates and all loss terms (death rate and other differentiation rates). We will later disentangle these two latter rates in order to refine our idea of the idea HSC activity (Section 3.2).

2.1.5   Number of active HSCs

So far we have determined how often an HSC proliferates or differentiates on average, but average rates are not informative on the actual number of HSCs that actively take part into haematopoiesis during the observation time. We define as active a cell that produces at least one differentiated progeny, but it is clear that the differentiation activity is, at least at the level of the cellular population, connected to the proliferation activity, in order to balance the number of cells.

To address this question, our experimental partners administrated low doses of tamoxifen to a cohort of mice and checked for the label in the mature blood cells. The idea was that, if the number of initially labelled stem cells is very low, the probability of having labelled an active stem cells should be lower too, and thus it might be likely to detect no labelled progeny in the analysed mice. However, this was not the case. Out of 61 mice with labelling frequency lower than 1%, and independently of the time between labelling and measuring, all mice had labelled mature cells. Nevertheless, we can use the information on the number of measured mice compared to the initial labelling frequency to estimate a lower bound of active cells as described below.

Let us define:

- f_a as the fraction of active HSCs (considered the same for all mice but unknown)

- N as the total number of HSCs (considered the same for all mice and known)

- L as the total number of labelled HSCs (variable among different mice)

We can now write the probability of labelling n_a active cells out of the total as:

$$P(n_a, f_a, L) = \frac{\binom{L}{n_a}\binom{N-Nf_a}{L-n_a}}{\binom{N}{L}} \tag{2.23}$$

and thus the probability of labelling no active cells:

$$P(0, f_a, L) = \frac{\binom{N-Nf_a}{L}}{\binom{N}{L}} \tag{2.24}$$

The observed quantity is the HSC labelling frequency in each mouse, from which we can compute L if we know N, the number of HSCs in a mouse. We fix N=16800 after [Boggs, 1984].

In Figure 2.6 A, we see the measured labelling frequencies in an increasing order. The final estimation for the probability of labelling no active cells depends on which frequency we use to calculate L, and the higher the frequency, the higher the estimated f_a. Since we want to estimate the lower bound for the fraction of active cells, we should in principle use the lowest labelling frequency (mouse 1), but this would be not reliable due to the fact that only a small aliquot of HSCs was analysed, thus yielding a noisy estimate for L (which, for the lowest frequency, is of the order of a few cells). Therefore we pool the data. In Figure 2.6 B, we show the relative error on the average labelling frequency of as many mice as indicated on the x axis[3] (following the increasing order as in left panel).
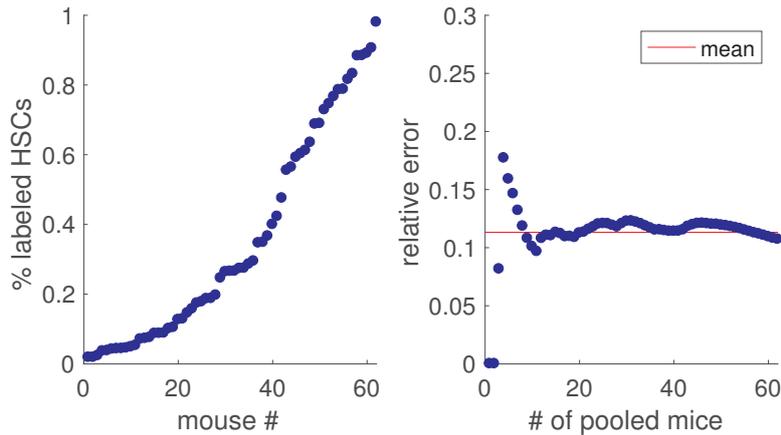
Figure 2.6: Limited dilution data. A: Labelling frequency for each mouse, plotted in increasing order. B: Relative error on the average of the labelling frequency taken on an increasing number of mice.

We see that after around 15 mice the error of the average value stabilises. We use this value for our calculation, namely an average labelling frequency of around 0.05%, which yields $L \sim 8$. We now have the probability of labelling no active cell only as a function of the fraction of active cells, $P_{0,L}(f_a)$. We

---

[3]The error corresponding to the first point from the left is 0 because it refers to just on mouse. It is 0 also on the following point because, by chance, the two lowest labelling frequencies happened to be equal.

use the binomial distribution:

$$B(m,k) = \binom{m}{k} (1 - P_{0,\mathrm{L}}(f_\mathrm{a}))^{m-k} \, P_{0,\mathrm{L}}(f_\mathrm{a})^k \qquad (2.25)$$

which gives the probability of finding k empty mice (i.e., no labelled progeny) out of m analysed. We use the information that all 15 mice we measured had labelled progeny, thus $m = 15$ and $k = 0$. We computed that the lowest f_a for which $B(15,0)$ is still significative($> 67\%$) is 0.3. In other words , if f_a were lower, it would have been very unlikely to not find any empty mouse out of the 15 measured. We can now state that at least around 5000 HSCs are active during the lifespan of a mouse.

Analogously, we were also able to determine the stem cell contribution to specific lineages of haematopoiesis:

- granulocytes (GR): 3.8% for granulocytes

- double positive lymphocytes (DP):0.4%

- precursors of bone marrow B cells (preB):0.7%

However, false negatives may occur which would enormously lower the estimated HSC contribution. False negatives may be due do under-sampling of differentiated cells appearing later than the time of measurement. Therefore, it os likely that the actual HSC contributions are greater than the estimated ones.

2.2   Activity of progeny

Analogously to what we have discussed so far for modelling the HSC and the ST-HSC compartments, we could in principle model all haematopoietic populations as defined by phenotypic markers.

Two caveats should, however, be taken into account. First, as we have pointed out earlier, the assumed lineage topology affects the estimation of efflux rates (see also Section 2.3). Second, if cells reside in different locations in the body, it is difficult to quantify the cell number ratio between the populations in question and their progenitors. In fact, we nevertheless attempted this quantification by analysing data from the major reasons (bone marrow, spleen, thymus) and estimating which fraction the aliquot analysed by FACS represented of the total organ. Specifically for the bone marrow, we relied on the estimation of HSCs as in [Boggs, 1984], and consider this to be a constant among the mice.

We modelled the kinetics of 16 populations according to the classical model

of haematopoiesis (Figure 2.7). Some populations (spleen populations other than granulocytes, megakaryocytes, macrophages) were not considered due to difficulties in isolation, lack of knowledge about their position in the tree or lack of nucleus and thus impossibility in distinguishing labelled from unlabelled cells (final stages of erythropoiesis). All cells were taken from the bone marrow, apart from the spleen granulocytes and the thymic T cells and precursors.

In the classical model, the most immature stem cell population, the HSCs, reside at the top of the developmental cascade and constantly supply cells to the direct progeny, the short term haematopoietic stem cells, ST-HSCs. ST-HSCs feed cells into the multipotent progenitors, MPPs. MPPs give rise to the common myeloid progenitors, CMPs [Akashi et al., 2000], and the common lymphoid progenitors, CLPs [Kondo et al., 1997]. On the one hand CMPs feed into MEPs, the megakaryocyte-erythroid progenitors, and into GMPs, the precursors of monocytes and granulocytes. On the other hand CLPs feed into the B cells and T cell development in the thymus (Figure 2.7).

To describe the equations used for modelling, we think of each population in turn as the reference, ref, compartment. Its direct progenitor is the upstream, up, compartment and its progenies are the downstream, $down$, compartments:

- stem cells

up	HSC	ST
$\boldsymbol{ref}$	**ST**	**MPP**
$down$	MPP	CMP,CLP

- myeloid cells

up	MPP	CMP	CMP	GMP	MEP	pro Ery
$\boldsymbol{ref}$	**CMP**	**GMP**	**MEP**	**GR s**	**pro Ery**	**baso**
$down$	GMP,MEP	GR	pro Ery	—	baso	—

- lymphoid cells

up	MPP	CLP	pro B	CLP	DN	DP	DP
$\boldsymbol{ref}$	**CLP**	**pro B**	**B**	**DN**	**DP**	**CD4**	**CD8**
$down$	pro B, DN	B	—	DP	CD4, CD8	—	—
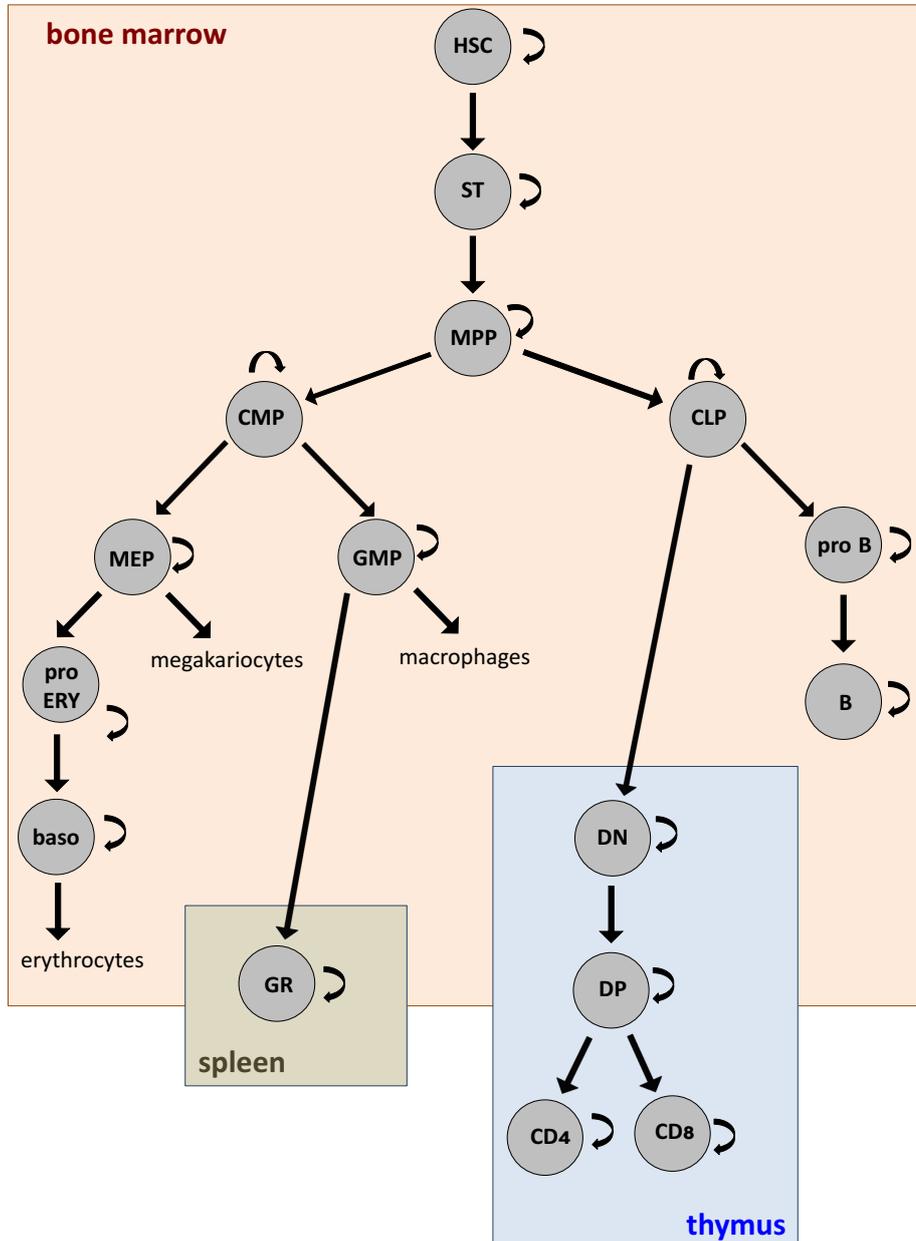
Figure 2.7: The classical scheme of haematopoiesis. In the figure we include with the circle (symbolising a cell) and the respective abbreviation all the populations for which data was collected. Other populations are included with the extended name. Background colour code refers to the different organs from which the cells were taken (bone marrow, spleen, thymus).

Reasoning as in the previous section, we model the relative increase in labelling frequency as:

$$\dot{f}_{ref}(t) = \kappa_{ref} \ (f_{up}(t) - f_{ref}(t)) \tag{2.26}$$

and the compartment size ratios as:

$$\frac{N_{ref}}{N_{up}} = \frac{\alpha_{up \to ref}}{\kappa_{ref}} \tag{2.27}$$

where $\alpha_{up \to ref}$ is now the differentiation rate from the upstream to a downstream compartment. Also, we define the net efflux as before:

$$\kappa_{ref} = \sum_{down} \alpha_{ref \to down} + \delta_{ref} - \lambda_{ref} \tag{2.28}$$

and introduce the concept of net proliferation, that is, the difference between the loss rate (differentiation plus death) and the proliferation rate:

$$\beta_{ref} = \lambda_{ref} - \delta_{ref} \tag{2.29}$$

This latter definition is useful to have an idea on the minimum value of the proliferation rate. The information contained in the label propagation and in the ratio of compartment size can be modelled upon using any combinations of parameters: net proliferation rate, differentiation rate and net efflux rate. We performed the modelling twice upon changing parametrisation in order to have direct estimate of all parameters and their profile likelihood confidence intervals. The model reproduces the data, as shown in Figure 2.8.

In Figures 2.10-2.12 we present all estimated parameters, best fit and profile likelihood confidence bounds.

The estimated parameters have several implications. First of all, we notice a decreasing residence time with increasing commitment along a lineage. This means that cells spend less time in a compartment as they mature, which implies that more immature cells act as a reservoir for haematopoietic cells while the more committed populations are transient compartments. Since their net proliferation is higher with respect to stem cells, we conclude that mature populations are rather amplifying compartments.

Along the myeloid linage, the upper bounds of all compartments net effluxes upper bounds were estimated to be 4 days, the maximum value allowed for running the optimisation. This is most likely a computational limitation: in the myeloid lineage the flux is so fast that equilibration is achieved instantaneously. More reliable is thus the lower bound on the net efflux which
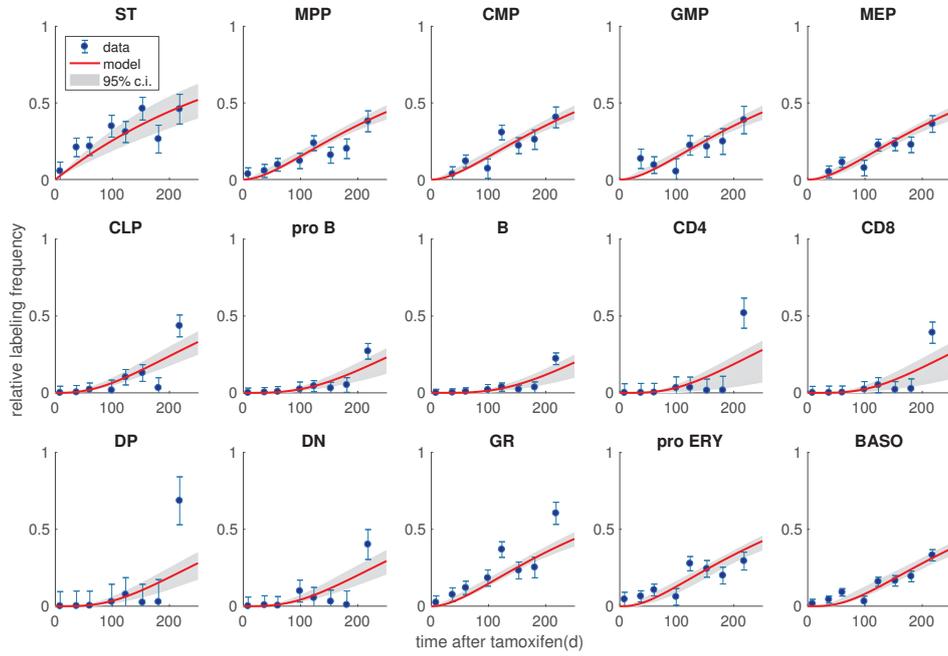
Figure 2.8: Fit of the classical model of haematopoiesis to the progression of fate mapping label introduced into HSCs. Grey areas: bootstrap confidence bounds on the model.
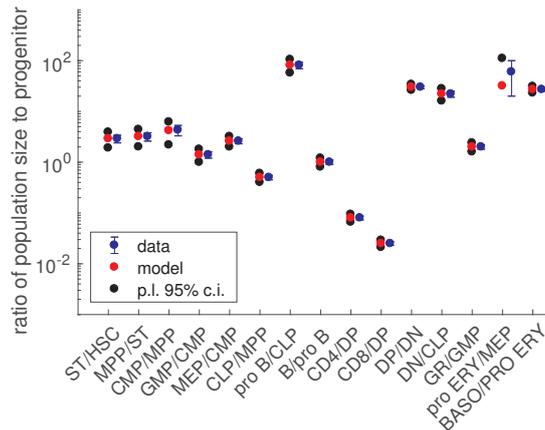
Figure 2.9: Fit of the classical model of haematopoiesis to the ratio of compartments. In grey: profile likelihood confidence intervals on the parameters. Lower bound for pro Ery/MEP is not shown because it is small (see computed profile likelihoods, Figure 8.4 in the Supplementary Material).
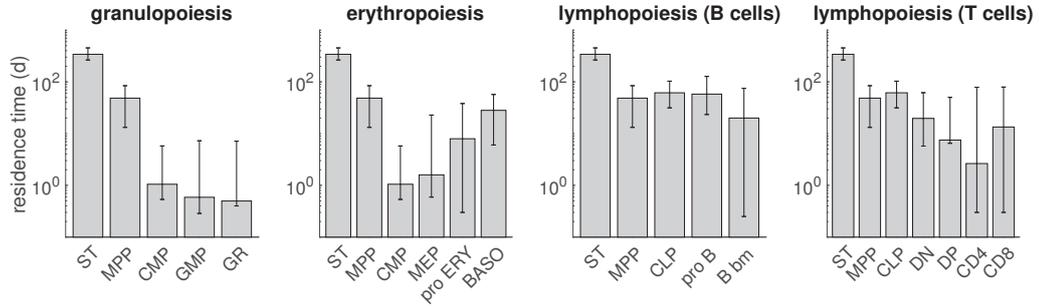
Figure 2.10: Best fit and profile likelihood 95% confidence intervals of the residence times for all lineages.
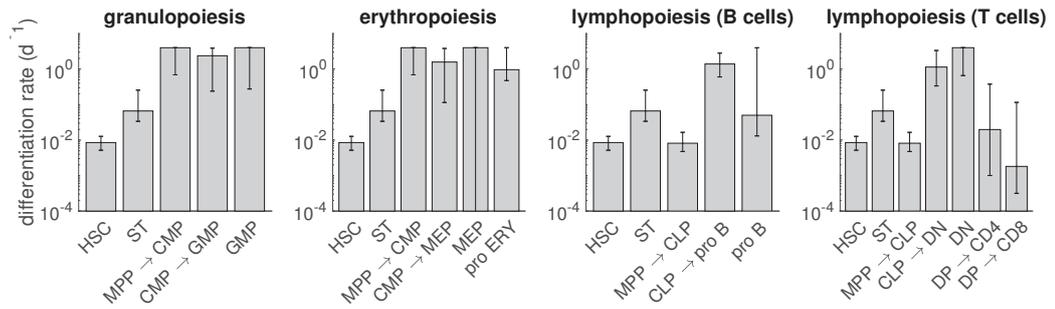
Figure 2.11: Best fit and profile likelihood 95% confidence intervals of the differentiation rates for all lineages.
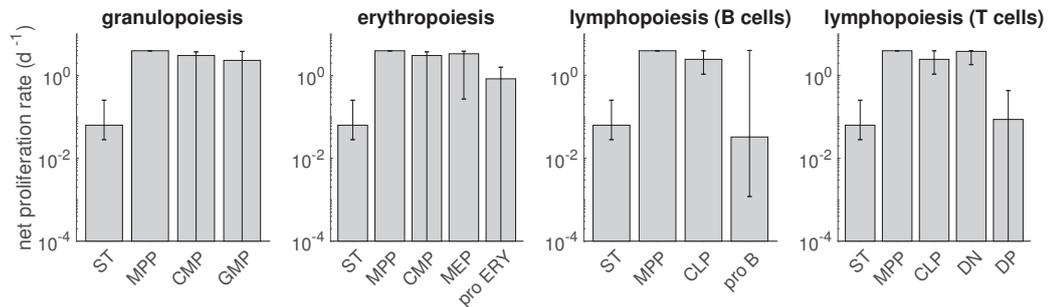
Figure 2.12: Best fit and profile likelihood 95% confidence intervals of the net proliferation rates for all lineages.

is one order of magnitude greater than MPP's. The only exception could be the basophils, precursors of the erythrocytes. This could be a biological relevant feature (long lived reservoir for the essential red blood cells), or an artefact due to the wrong assumed topology (several studies have suggested an earlier erythroid commitment, see Section 2.3).

In the lymphoid lineage the common progenitors and the progenitors of B cells are long lived, and so are, potentially, cells in the thymus, in agreement with the suggested self-renewal of the thymus [Martins et al., 2012]. Once again, however, parameters often lack an upper bound other than the prescribed upper bound of 4 per day. To summarise our finding, we plot the time courses of label increase in Figure 2.13. We see that all lymphoid cells lag behind the myeloid cells, which are indistinguishable from MPPs. ST-HSCs are significantly distinct from the others because their slow kinetics generate a delay until labelled cells reach downstream cells populations.
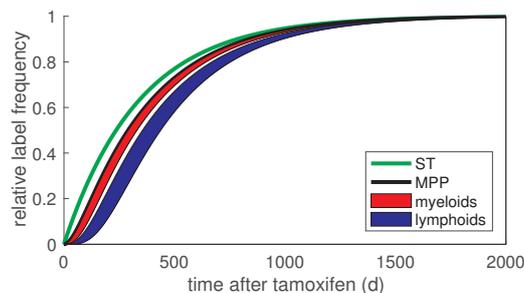
Figure 2.13: Best fit and 95% bootstrap confidence intervals of the time-dependant equilibration of the fraction of labelled cells in all compartments to HSC.

Differentiation rates also tend to increase along lineages, with few exceptions, and the total flux of cells also increases (Figure 2.15). Finally, net proliferation rates also tend to increase, although in some cases they are not identifiable.

We then compared our prediction for the time evolution of the labelling frequency with data measured up to two years after tamoxifen induction (data measured by Dr. Katrin Busch, unpublished). For myeloid compartments the model prediction is largely corroborated. However, the label accumulation in the lymphoid lineages is generally overestimated. This could indicate that lymphoid output from HSCs declines with ageing. Moreover, the steady state assumption made in the model might no longer be valid for older mice. In particular, the decrease (or slower increase) of the label frequency in a reference compartment could be due to a smaller differentiation
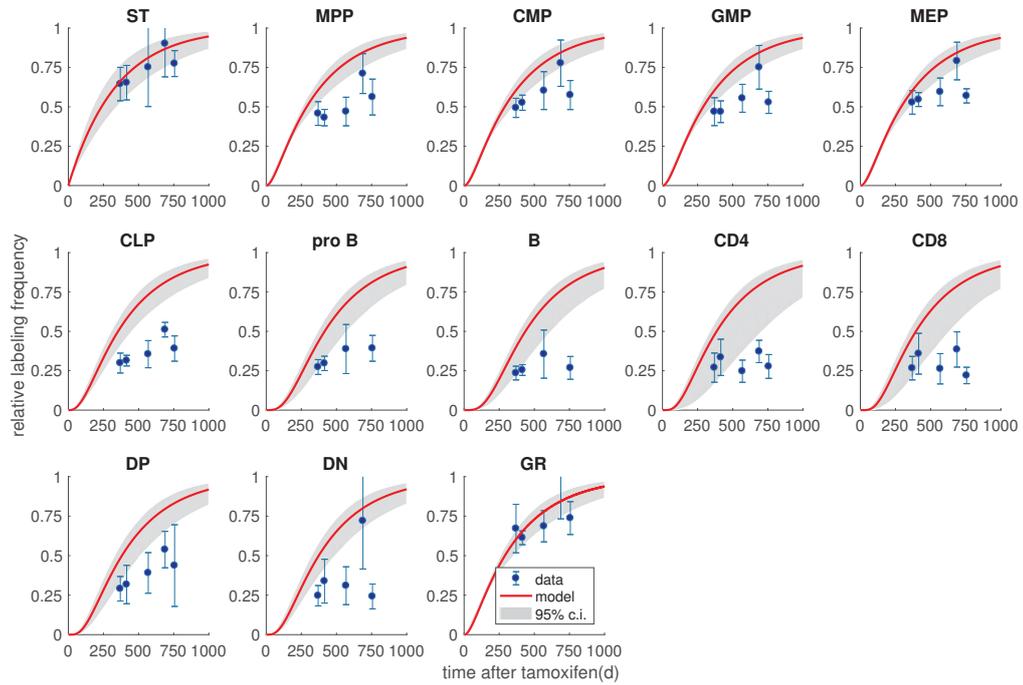
Figure 2.14: Best fit and profile likelihood 95% confidence intervals of the prediction for the label frequency in old mice.

rate from the upstream compartment or of the reference compartment itself, which would imply a decrease, at least temporarily, of haematopoietic output. To investigate this further, we will consider a model for non-stationary haematopoiesis in Chapter 5.

Finally, upon reparametrising the model one can also obtain confidence intervals on the population size ratios relative to HSCs and consequently, knowing the number of HSCs in a mouse, infer the size of a population:
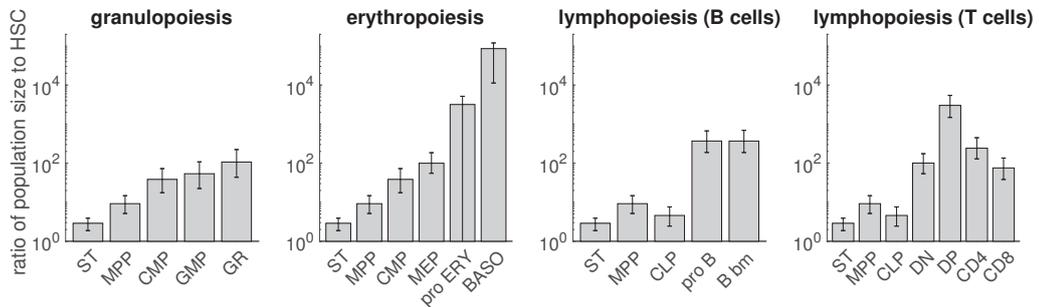
Figure 2.15: Best fit and profile likelihood 95% confidence intervals of the ratio of compartment sizes to HSC for all lineages.

27
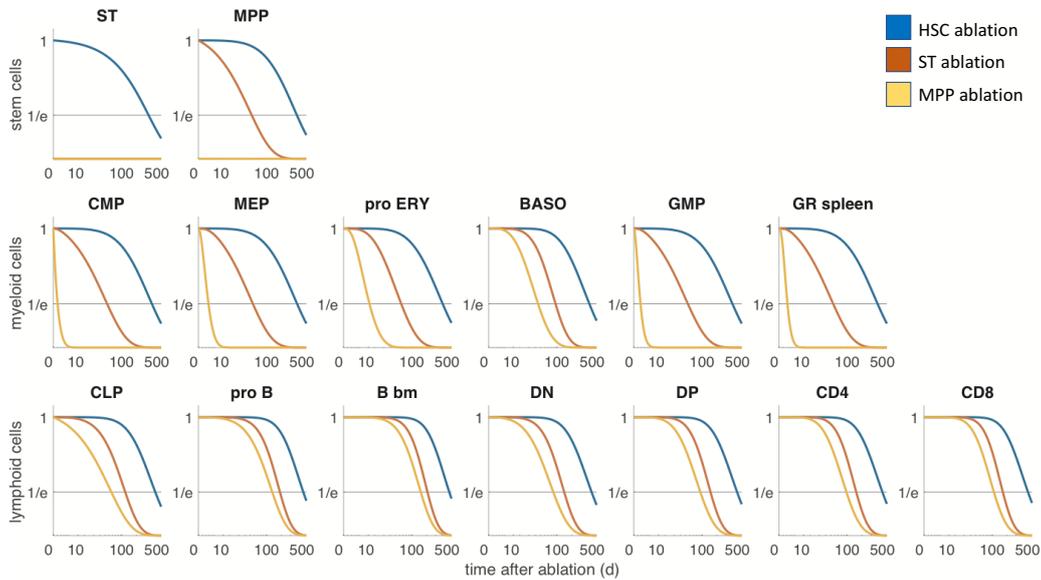
Figure 2.16: Simulation of the instantaneous ablation of HSCs, ST-HSCs or MPPs. The number of cells in the downstream compartments drops with different time scales.

As we did in the previous section, we again investigate the effects of the ablation of stem cells on the progeny, assuming that no feedback mechanism is triggered by the ablation. In Figure 2.16 we show how the number of cells drops from the steady state value (normalised to 1 for each population) when either HSCs, ST-HSCs or MPPs are instantaneously ablated. We had already noticed that it takes around one year to have a significant drop in cell numbers after HSC ablation due to the long residence time of ST-HSCs. On the other hand, if ST-HSCs themselves were ablated, we would see effects already after 70 days due to the shorter residence time of MPPs. Finally, MPPs ablation would cause an immediate effect on the myeloid lineage, but still a retarded effect on the lymphoid lineage, due to CLPs' high residence time, and an even more delayed effect on the thymus due to the the residence time in the double negative cells compartment, DN.

2.2.1 Branching points

The study of the full scheme for haematopoiesis also poses new questions regarding lineage topology. This concerns, for example, the position of a branching point. In the classical scheme we are been examining so far, we considered four branching points:

- MPPs splitting into CMPs and CLPs

- CMPs into MEPs and GMPs

- CLPs into pro Bs and DNs

- DPs into CD4 and CD8 T cells

For these branching points we now quantify the ratio of cell fluxes towards the different fates.

In order to estimate the flux ratio, we reparametrise our model in such a way that it contains flux ratio at each branching point as an explicit parameter. Fitting this model to the data yields the flux ratio with confidence bounds (Figure 2.2.1). While the fluxes from CMPs and DPs cannot be identified,
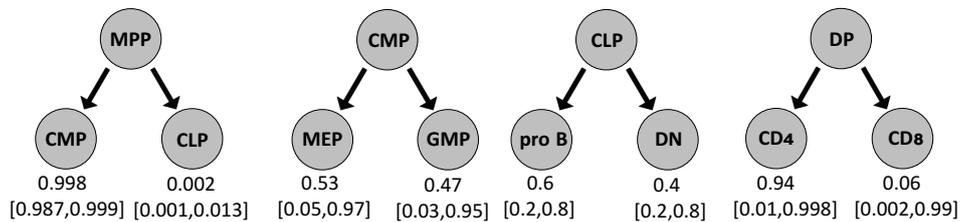
Figure 2.17: Best fit and profile likelihood confidence bounds (95%) flux ratios at branching points.

this analysis shows that the vast majority of MPPs become myeloid cells and at least 20% of CLPs become thymocytes.

Finally, we compare the position of the myeloid versus lymphoid branching point in the classical scheme to an hypothetical earlier branching, after the ST-HSC compartment, with MPPs being now the progenitor of the solely myeloid cells, and CLPs arising directly from ST-HSCs, as suggested by [Oguro et al., 2013] (Figure 2.18). The classical scheme suggests a flux into myeloid cells being at least 2 orders of magnitude greater than the flux into lymphoid cells. In the alternative scheme, the ratio of fluxes can go down to 3 fold, without changing the estimation for the total number of cells that become common myeloid progenitors per unit time. This is possible via selective proliferation in the MPP compartment.

29

a) Classical scheme

b) Alternative scheme
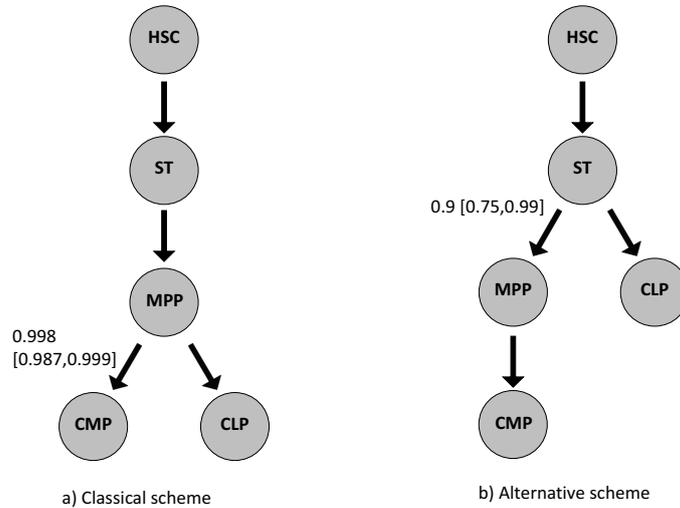
Figure 2.18: Possible schemes for the position of the myeloid versus lymphoid branching point. Best fit and profile likelihood estimation for the fraction of flux into the myeloid cells are shown.

This type of analysis is an example of the different interpretation of the results obtained considering different topologies: in the classical view, the large flux flowing into myeloid cells is achieved at the level of MPP through a differentiation bias. In the alternative view, most of the myeloid flux is due to selective proliferation of myeloid-biased precursors.

2.3 Lineage topology

In order to determine the kinetic properties of haematopoietic stem and progenitors cells we have relied in the previous section of the classical model of haematopoietic differentiation pathways. Several studies suggested alternative routes, based on transplantation and inference from single-cell transcriptome data [Nutt et al., 2005] [Lai and Kondo, 2006] [Guo et al., 2013][Yamamoto et al., 2013] [Perié et al., 2014].

Here, we study the following question: Do population fate mapping data contain information on lineage topology? And in turn: How do parameter estimates change when one considers different developmental pathways?

2.3.1 The progenitor-progeny relationship

In the fate mapping data described in Section 2.1 a genetic label is switched on in stem cells and progressively propagated to the progeny. Hence progeny

should always have a lower labelling frequency than its progenitor cells. Is the classical differentiation scheme compatible with the data in this respect?

To address this point we compare the labelling frequencies of all pairs of putative progenitor and progeny over time. We recall that the data consists in labelling frequencies measured in sometimes small aliquots, which poses the problem of sampling noise, especially at early time points when progeny labelling frequencies are small. To smoothen outliers effects we computed the moving average of the ratio of labelling frequency of the progeny with respect to the putative progenitor over time. For example, ST-HSCs are thought to be the direct progeny of HSCs. All other cell types might emerge from ST-HSCs without the need of a direct flux from HSCs. If this is true, then the ratio of label frequency in any compartment with respect to the labelling ratio in ST-HSCs should never exceed one. This is indeed the case (Figure 2.19).
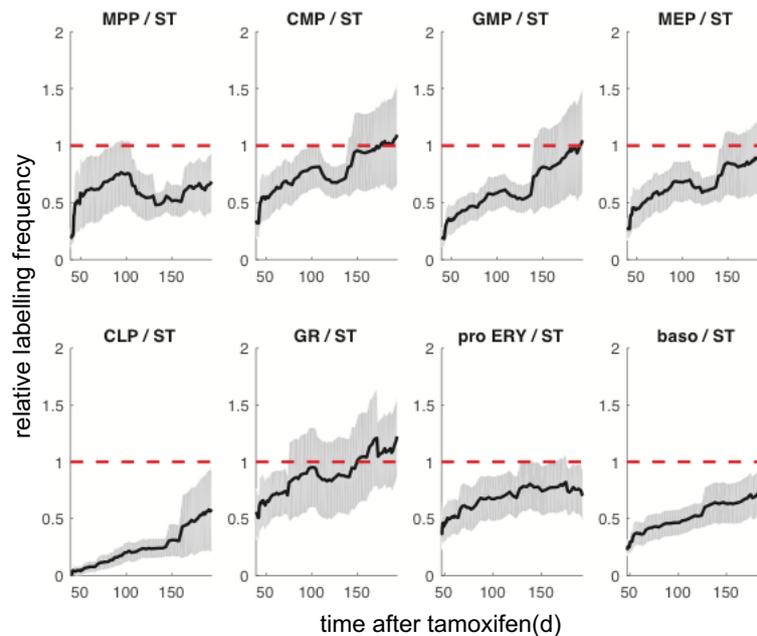
Figure 2.19: Moving average with 95% confidence intervals of the ratio of labelled cells in progeny compartments compared to ST-HSCs. Since no ratio significantly exceeds 1 over the observed time, there is no need to assume direct contributions from HSC to the more differentiated compartments.

Of note, the ratio being lower than one is only a necessary, not a sufficient condition for HSC not being a/the direct progenitor of committed cells. Other quantitative approaches could shed light on this issue (see next

section).

We repeat the analysis computing the moving average this time for the ratios with respect to MPP (Figure 2.20).
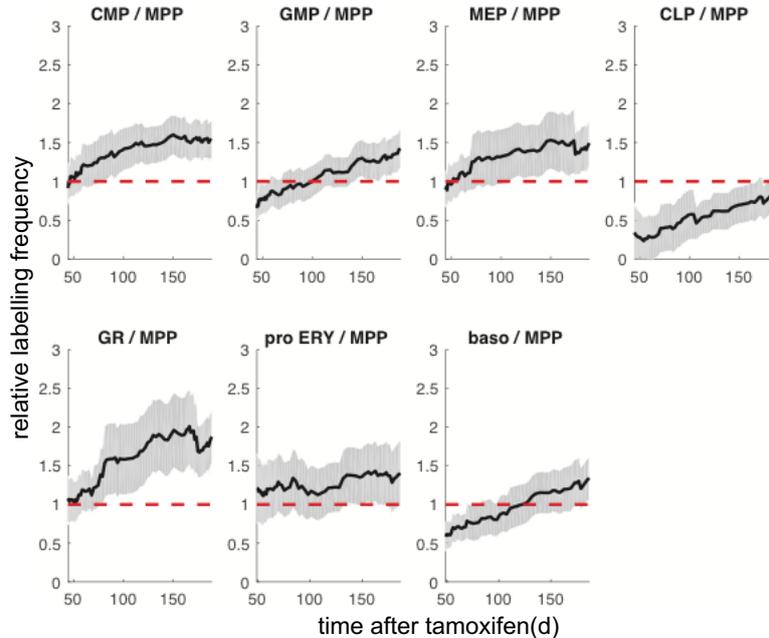
Figure 2.20: Moving average with 95% confidence intervals of the ratio of labelled cells in progeny compartments compared to MPPs. The ratios for CMPs and GRs significantly exceed 1, suggesting the existence of additional pathways into these populations that bypass MPPs or heterogeneity in the MPP compartment with respect to myeloid fate.

In this case, there is a time dependant increase in the ratios of labelling frequencies in the myeloid cells respect to that of MPPs. This might hint at a bypass from ST-HSCs directly into the myeloid populations, or at least into CMPs and GRs, for which the ratio significantly exceeds 1.

Finally, another kind of information can be contained in the labelling progression data. HSC is the only initially tagged population and we observe the label emerging in all other populations along time. Thus HSCs must be the founders of haematopoiesis. But if HSCs were heterogeneous, with different sub-populations feeding the downstream compartments, one could observe the labelling frequency in the progeny overshooting that of HSCs at later time points (see Figure 8.5 in the Supplementary Material). This was not the case as shown in Figure 2.21.
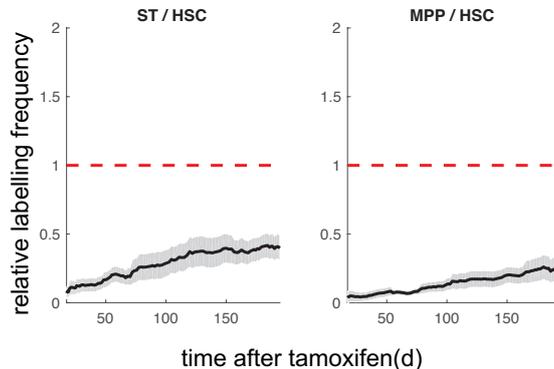
Figure 2.21: Moving average with 95% confidence intervals of the ratio of labelled cells in progeny compartments compared to HSC. The trajectories are entirely below one, as expected for a homogeneous progenitor population.

2.3.2 Inferring topology

From the analysis developed in the previous section, it emerged that myeloid cells might be fed directly from ST-HCSs, and possibly additional pathways from MPPs and myeloid progenitors might exist.

To examine this further we fit a model that has all possible differentiation pathways from all compartments to all progenies. For simplicity we only consider myeloid lineages and keep the order of the populations, so that MPPs are still upstream of GMPs and can differentiate directly into it, but not the other way round. Backward differentiation is also excluded has it has not been observed in transplant experiments. We then compared this model to classical model, and to three further models in which all myeloid cells directly come from HSCs, ST-HSCs or MPPs respectively. We then ranked the models according to the corrected Akaike information criterion AICc (see Methods), as shown in Table 2.1:

Model	# data	# parameters	χ^2	$AICc$
differentiation from HSCs	60	14	67	104
differentiation from ST-HSCs	60	14	70	107
classical	60	14	81	118
differentiation from MPPs	60	14	85	122
full scheme	60	29	60	176

Table 2.1: Ranking of models that differ in their topology.

We thus see that the most informative models that still produce a good fit are those in which progeny is produced from the very top of the hierarchy, by either HSCs or ST-HSCs. Since the difference of their AICc is less than ten they are indistinguishable, and with similar reasoning the other models are significantly less informative. In Figure 2.22 we show that, upon assuming different lineage topologies, estimated parameters, for example the residence times, change. We see that for CMPs, GMPs and GRs the estimated efflux
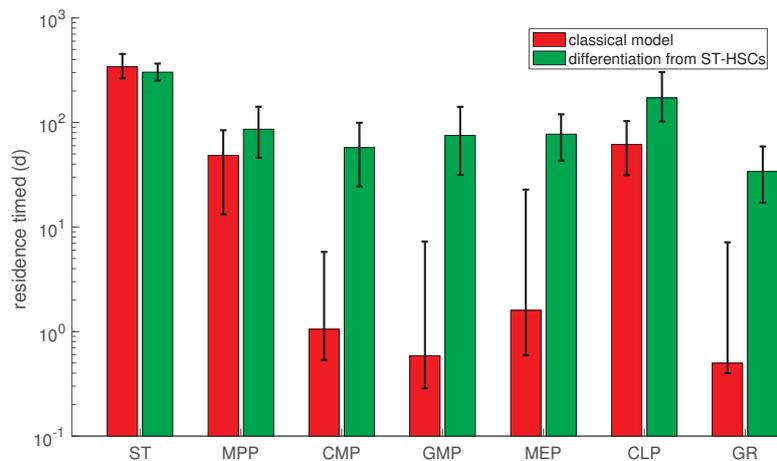
Figure 2.22: Comparison of the estimated residence time for the classical model and an alternative model where all populations downstream of ST-HSCs are generated from ST-HSCs directly.

rates are different by an order of magnitude. Changing topology leads to different conclusions on the kinetic parameters that are to be estimated.

Chapter 3

Consideration of compartment heterogeneity and cell proliferation

In the previous chapter we saw how the fate mapping data, combined with the measurements of compartment sizes, allow to estimate the residence times and the differentiation rates of the populations of interest. The data also provides information on the net proliferation rate, (the proliferation rate minus the death rate). In order to disentangle these latter rates, we need additional data and related to cell cycle or apoptosis measurements. We used previously published data on BrdU accumulation in haematopoietic stem and compartment cells[Oguro et al., 2013] whose features we describe in the following before introducing the modelling approach.

3.1 Haematopoietic populations revisited

[Oguro et al., 2013] found new SLAM markers that stratify the most immature haematopoietic cell populations. Figure 3.1 shows such populations and also compares them to our definitions as published in [Busch et al., 2015].

The differentiation hierarchy proposed by [Oguro et al., 2013] is established based on reconstitution potential after primary and secondary transplantation, self-renewal, multipotency, fraction of active cells and frequency of proliferation, that is, on accepted definitions on "stemness". This does not necessarily imply that a cell with more 'stemness" gives rise to a cell with less 'stemness", although this is compatible with intuition, since it is clear that the most mature, functional cells have no reconstitution potential, cannot self-renew and so on.
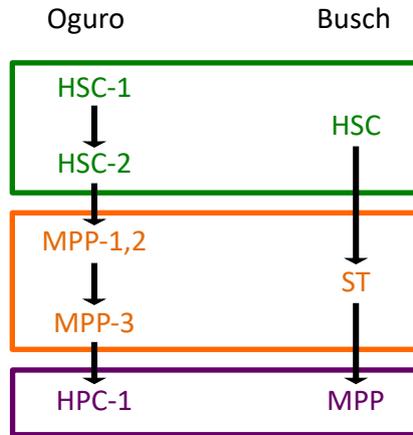
Figure 3.1: A proposed more detailed definition of haematopoietic stem and progenitor cells

Since for modelling purposes it is fundamental to know the correct order of differentiation (see Section 2.3), we first checked whether the suggested order is compatible with the label propagation experiment, or, in other words, whether a putative progeny compartment has a lower fraction of labelled cells with respect to its putative progenitor[1]. Figure 3.2 shows that, at different time points the frequency of YFP$^+$ cells is compatible with Oguro's proposed hierarchy, which will be therefore consider correct from now on.

Note that MPP-1,2 corresponds to the MPP-1 and MPP-2 populations taken together, since MPP-1 is a very small population that gives rise to noisy data. Such data is nevertheless not contradicting the differentiation hierarchy.

3.2   Cell cycle models

[Oguro et al., 2013] confirmed the haematopoietic hierarchy by means of the measurement of the fraction of BrdU$^+$ cells over time. BrdU is a thymidine analogue that is incorporated in the genome upon DNA replication, thus during the S phase of the cell cycle. In the presence of BrdU the label per cell keeps increasing.

In our case BrdU is administrated continuously via drinking water, thus we assume that the label is not diluted. Moreover, as more and more cells go through cell cycle, and thus through S phase, the fraction of BrdU$^+$ cells

---

[1]Unlike the case of Section 2.3, we do not have enough data here to repeat the moving average analysis. Plus, that was necessary to show whether an excess in the labelling ratio in a progeny with respect to a progenitor is significant, while here we see that there is no excess at any time point
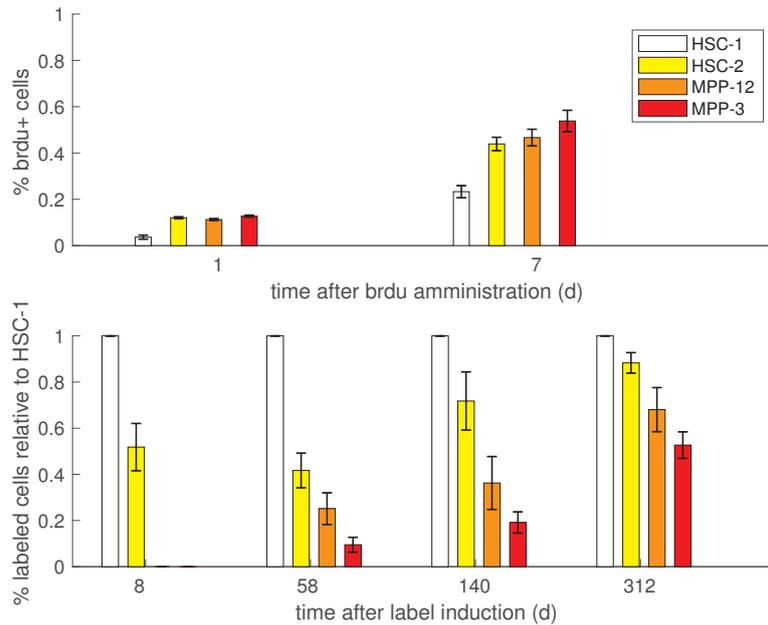
Figure 3.2: Label propagation in the refined populations. Since no labelling frequency in a progeny exceeds that of its putative progenitor, there is no need to question the suggested hierarchy.

over time increases and approaches the fraction of proliferatively active cells within a population for a long enough time.

Here the concept of an active cell is even more subtle than what we mentioned in Section 2.1.5. There, we estimated the fraction of cells that are active over the period of observation. Here, a suitable model can estimate the fraction of cycling cells even for short observation times, when not all cycling cells have completed a cycle. This will become clearer later in the section.

Since one of the cell cycle phases, namely S, is special due to the mechanism of accumulation of BrdU label, one should in principle build a model that describes the progression through the cell cycle (instead of just assuming a cell proliferation rate). Figure 3.3 shows the classical view of the cell cycle, where a cell progresses from G_1 to S, then G_2, M and then again G_1 phase. The cells may also spend time in a quiescent state, G_0, derived from G_1.

The minimal model to account for the accumulation of BrdU label would be a model in which the S phase has a special role, since during the S phase the cells becomes BrdU positive. This could be achieved upon merging the
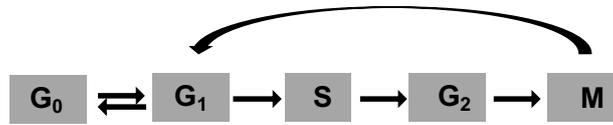
Figure 3.3: Scheme of the cell cycle. Arrows represent phase progression.

G_0 and G_1 phases, the G_2 and M phases, and assuming that $BrdU^-$ cells in S phase become $BrdU^+$ cells in G_2/M phase (Figure 3.4).
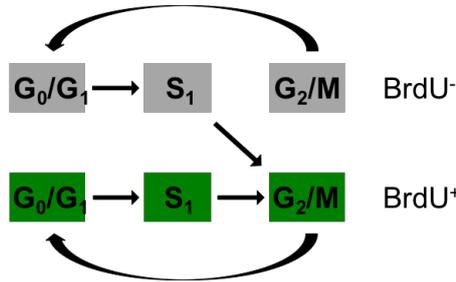
Figure 3.4: Simple model for BrdU accumulation. A cell becomes $BrdU^+$ in the S phase and keeps on cycling, or becomes quiescent, remaining positive. Straight arrows represent phase progression, curved arrows proliferation. Cell death is not shown.

This model can be further improves upon splitting the S phase into n sub-phases, $\dfrac{1}{2\,n}$ being the fraction of DNA that a cell must synthesise in order to become $BrdU^+$ for the detector (the 2 comes from the fact that only one strand is newly synthesised). Cells exiting any of these sub-phases directly go to the next sub-phase or to the G_2/M phase if they were in S_n and become, or remain, positive (Figure 3.5).

Now, we aim at replacing this model with an even simpler model in which a cell is seen as a black box and the phases are not resolved, as we have been doing for the previous analysis. If this were possible, it would have two advantages: removing the arbitrariness in the choice of n and reducing the number of parameters, and equations, thus speeding up computation.

The idea is that, if the number of cells in either S, G_2 or M phases is small compared to the number of cells in either G_0 or G_1, then we can say that the near majority of cells that are BrdU positive have gone through at least one division (positive cells in either G_0 or G_1). In this way, one would interpret BrdU positive cells as the cells that have gone through at least one division.
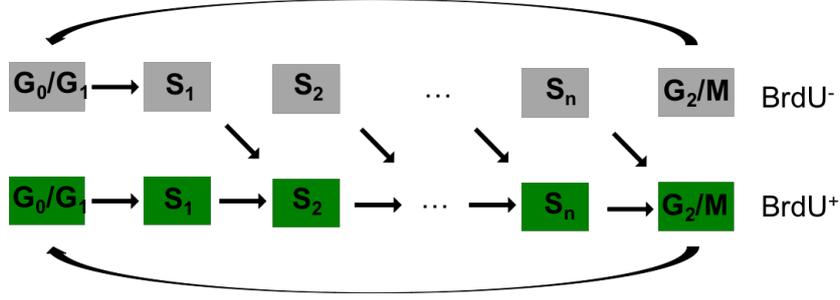
Figure 3.5: More detailed model for BrdU accumulation. A cell becomes BrdU positive if it has synthesised enough DNA, which is represented as a cell exiting an S sub-phase, and then keeps on cycling, or becomes quiescent, remaining positive. Straight arrows represent phase progression, curved arrows proliferation. Cell death is not shown.

In Section 2.1 we defined all the parameters that correspond to the possible cell fates. Based on that we now write equations that describe the BrdU accumulation in the cells. In the following, $N_A^{+/-}(t)$ are the numbers of BrdU positive/negative cells at time t in a stem-cell-like population A. After one division, negative cells become positive and stay positive, but a symmetric division creates two positive cells of type A, an asymmetric division creates two positive cells, one of type A and one of type B, one symmetric differentiation creates two positive cells of type B. One-to-one differentiation of a negative/positive cell of type A gives rise to a negative/positive cell of type B, and death just removes the cell. Figure 3.6 shows all the possibilities and compares them to the outcomes of label propagation in fate mapping.

The equations for the number of BrdU negative and positive cells are:

$$\dot{N_A^-}(t) = - (\sigma_A + \rho_A + \mu_A + \gamma_A + \delta_A) \, N_A^-(t) = \tag{3.1}$$
$$- (2 \, \sigma_A + \gamma_A) \, N_A^-(t) \tag{3.2}$$
$$\dot{N_A^+}(t) = (2 \, \sigma_A + \gamma_A) \, N_A^-(t) + (\sigma_A - \rho_A - \mu_A - \delta_A) \, N_A^+(t) = \tag{3.3}$$
$$(2 \, \sigma_A + \gamma_A) \, N_A^-(t) \tag{3.4}$$

where the RHS of the equations has been obtained via the steady state assumption: $\sigma_A - \rho_A - \mu_A - \delta_A = 0$. Also, in steady state the total number of cells is constant, thus the equations are not independent and we pick just one of them for modelling purposes. Since the observable is the fraction of BrdU positive cells, $B_A^+(t)$, we finally write:

$$\dot{B_A^+}(t) = (2 \, \sigma_A + \gamma_A) \, (1 - B_A^+(t)) \tag{3.5}$$
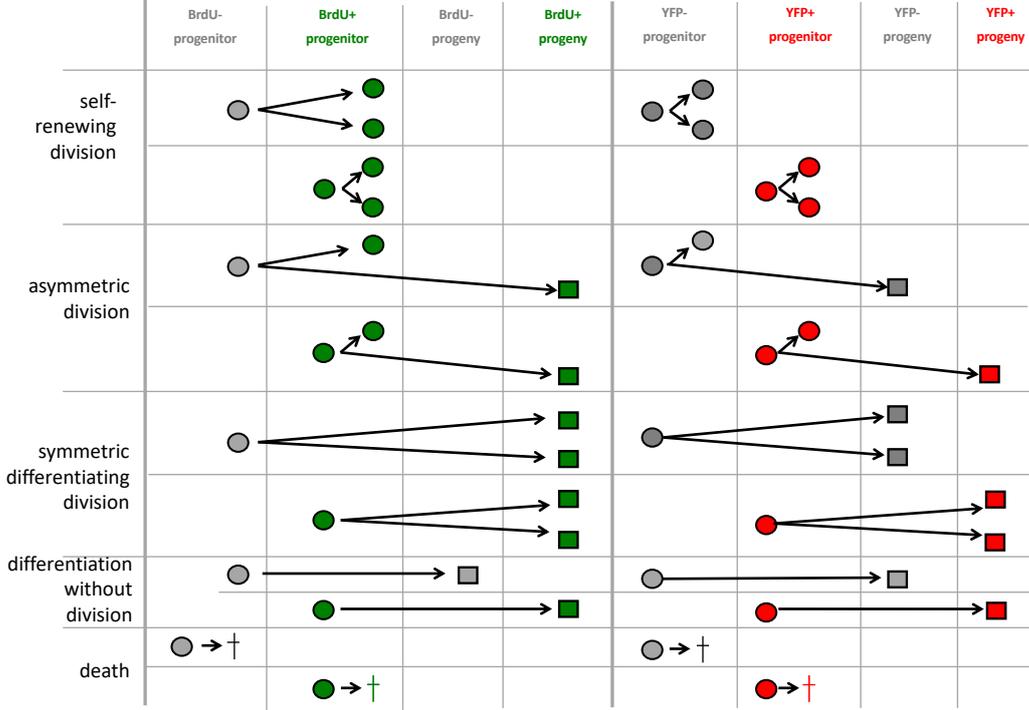
Figure 3.6: Possible fates for BrdU$^{+/-}$ and YFP$^{+/-}$ cells.

where we have used $B_A^-(t) + B_A^+(t) = 1$. In other words, the equations we just obtained should only work if all cells of type A were active. In reality, though, they still hold for slow kinetics and if the percentage of active cells is at least 30%, which is the case for our stem cells, as we showed in Section 2.1.5 (see Figure 8.6 in the Supplementary Material for more details).

We can reason analogously for the direct progeny of A, P:

$$\dot{B}_P^+(t) = (2\ \rho_A + \gamma_A + \mu_A\ B_A^+(t))\frac{N_A}{N_P} + (2\ \sigma_P + \gamma_P)\ (1 - B_P^+(t)) - \kappa_P\ B_P^+(t) \tag{3.6}$$

In our biological system, P represents any progeny of the stem cells, which are believed to be all active, thus again we assume $B_P^-(t) + B_P^+(t) = 1$.

We used this simplified model for BrdU accumulation to compare its performance versus the model in Figure 3.5. We simulated the phases and the black box model to reproduce the BrdU accumulation data taking into account the phases distribution as measured in [Oguro et al., 2013]. While for the most immature compartments there is hardly any difference in the estimated proliferation rates, for HPC-1,2 the black-box model yields a 2-fold higher proliferation rate than the more detailed model (Figure 8.7 in

40

the Supplementary Material). So we have to keep in mind that for those compartments we rather estimate an upper bound for the proliferation rate.

We now address the point of parameter identifiability. We use the method of Taylor series approach for linear systems [Pohjanpalo, 1978]. If $y_i(t, \mathbf{p})$ is the i-th observable as a function of the time and of the vector of parameters, then all its Taylor coefficients, $y_i^{(k)}(0, \mathbf{p})$, are unique for the measured output. We thus have as many linear combinations of identifiable parameters as independent linear combinations of Taylor coefficients. If our only observables were the BrdU data, we would have, for the system stem cell S and direct progeny, P, the following independent Taylor coefficients:

- a) $B_{\mathrm{S}}^+(0) = 0$

- b) $B_{\mathrm{S}}^{+(1)}(0) = 2\,\sigma_{\mathrm{A}} + \gamma_{\mathrm{A}} = S_1$

- c) $B_{\mathrm{P}}^+(0) = 0$

- d) $B_{\mathrm{P}}^{+(1)}(0) = \frac{(2\,\rho_{\mathrm{A}}+\gamma_{\mathrm{A}})\,\kappa_{\mathrm{P}}}{2\,\rho_{\mathrm{A}}+\gamma_{\mathrm{A}}+\mu_{\mathrm{S}}} + 2\,\sigma_{\mathrm{P}} + \gamma_{\mathrm{P}} = P_1$

- e) $B_{\mathrm{P}}^{+(2)}(0) = \frac{\mu_{\mathrm{S}}\,\kappa_{\mathrm{P}}\,S_1}{2\,\rho_{\mathrm{A}}+\gamma_{\mathrm{A}}+\mu_{\mathrm{S}}} + (2\,\sigma_{\mathrm{P}} + \gamma_{\mathrm{P}} - \kappa_{\mathrm{P}})P_1 = P_2$

from which hardly any interesting combination of parameters can be identified. Let us see what happens if we use the information coming from the fate mapping (κ identifiable) and the steady state ratio of compartments' size ($2\rho + \gamma + \mu$ identifiable). Note that items a) and c) do not give information. Upon considering all observables, we finally find that the following combinations of parameters can be in principle identified for each population:

- μ

- $2\,\sigma + \gamma$

- $2\,\rho + \gamma$

Note that already this would prove quite informative on the mechanism of differentiation (direct or upon proliferation). In addition, this identifiability analysis is not informative on practical identifiability, (i.e. the existence of lower or upper bound of the parameters). It might well be that we can practically estimate at least one bound for some parameters, as we will see in the following Section.

3.3 Estimation of proliferation rate

We thus modelled proliferation, fate mapping (for both the new data on refined populations and the previous data for more homogeneous populations,

this time accounting for underlying heterogeneity, see Figure 3.2) and compartment size ratios simultaneously with the models discussed above. One last simplification is to consider that only the most immature stem cells (HSC-1,2) divide other than symmetrically. All data could be modelled (Figure 3.7), and most parameters identified (Figure 3.8). Identifiability improved, as expected since we increased the number of data.
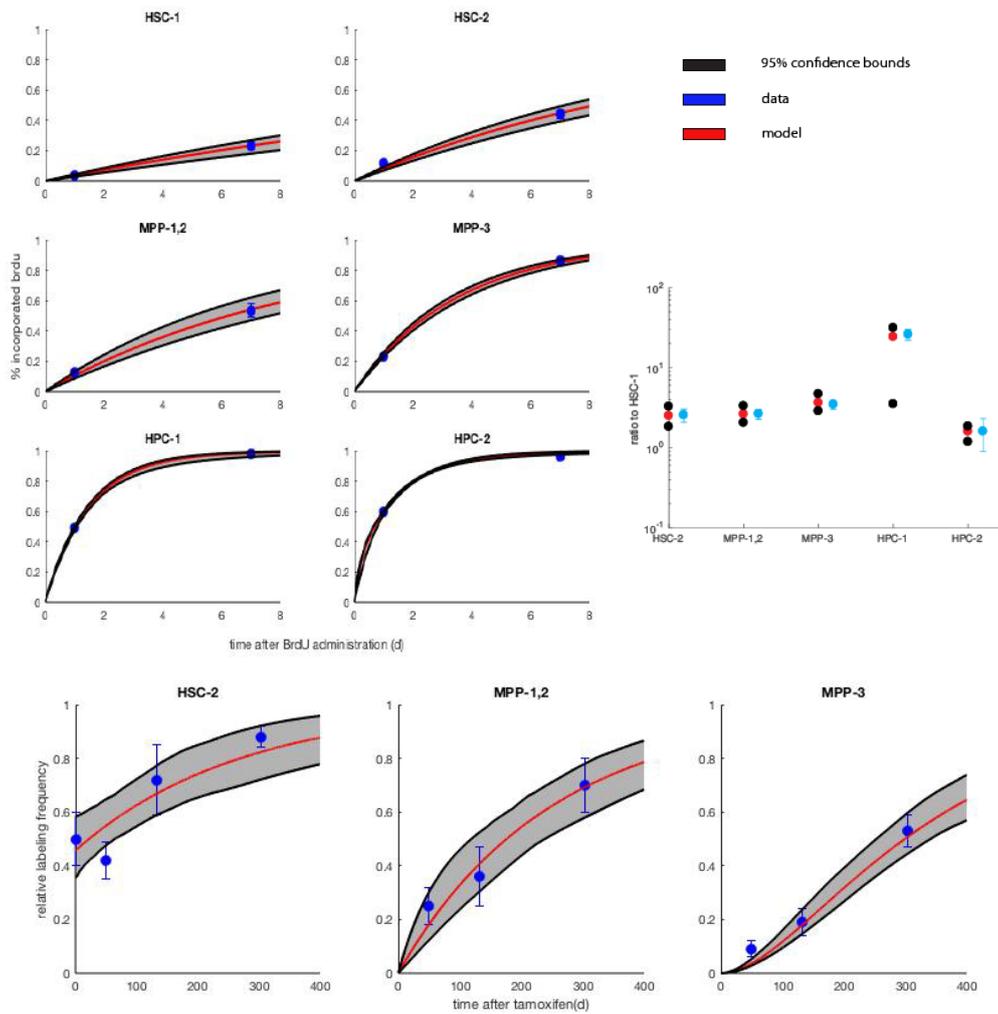
Figure 3.7: Model performance for the simultaneous fit of all sets of data.

Again, we confirm the trends we had already found in the case of the homogeneous model: increasing differentiation potential, decreasing residence time along the differentiation path. Moreover, we saw a clear trend for the proliferative potential, while for the death rate we had identifiability prob-
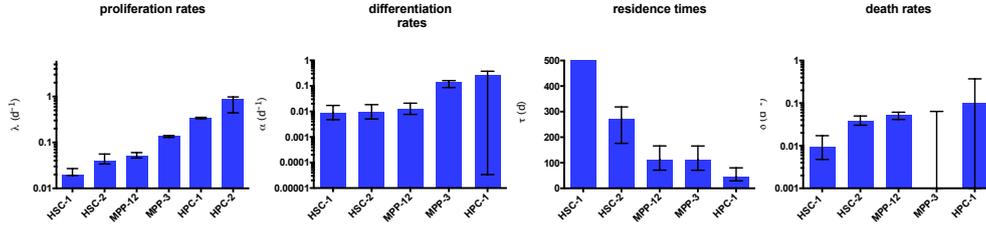
Figure 3.8: Best fit and profile likeliwood 95% confidence bounds for all model parameters.

lems, which only allows to state that the trend could be increasing.

Interestingly, we found an upper bound on $\rho + \gamma$ and a lower bound for $\sigma + \gamma$ for both HSC1 and HSC2, which resulted in a lower bound for the fraction of symmetric divisions (Figure 3.9 a). Estimates for the fraction of total symmetric divisions for the HSCs without considering the sub-compartments turn out to be even more strictly bounded, which implies that symmetric divisions must happen more frequently than the others and might also be the only mechanism of division (Figure 3.9 b).
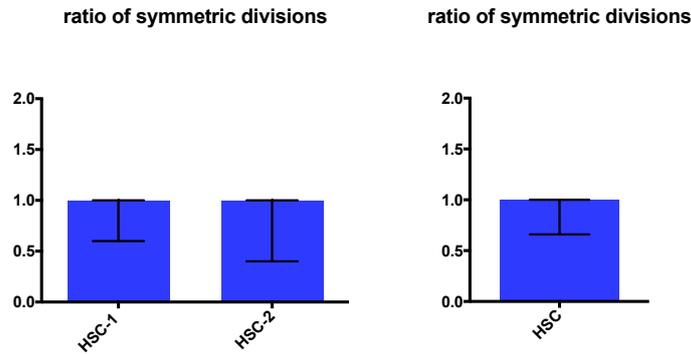
Figure 3.9: Best fit and profile likelihood for the ratio of symmetric divisions out of total divisions for a) HSC-1,2 and b) HSC.

Regarding differentiation, the data are not informative enough to determine the exact mechanism of commitment, but it is useful to notice several features. Asymmetric division cannot be the only mechanism to provide differentiated progeny and could also not be present. Symmetric or direct differentiation could each supply all the progeny. Nevertheless, we reckon that more data (thus smaller uncertainty in the measurements) could help distinguishing the two mechanisms.

3.4 Counting proliferations

It has been suggested [Bernitz et al., 2016] that stem cells divide symmetrically a fixed number of times before differentiating, instead of being able to stochastically differentiate or proliferate at any time point (with different probabilities). We want to test whether the data we dispose of contain the information on such a proliferation-counting mechanism. One can model a population B by writing equations for the number of cells of type $B_i(t)$ that have proliferated i times, assuming:

$$B_0(t) = \alpha_A A(t) - \sigma_B B_0(t) \tag{3.7}$$

$$B_i(t) = \sigma_B(2\ B_{i-1}(t) - B_i(t)) \qquad 1 \leq i \leq n \tag{3.8}$$

$$B_{n+1}(t) = 2\ \sigma_B\ B_n(t) - (\delta_B + \mu_B)B_{n+1}(t) \tag{3.9}$$

Note that:

- Population B has an influx from A, otherwise is not possible with this model to describe a steady-state scenario. Thus, we will not apply the model to the HSC-1 population.

- We assume that cells do not die in the proliferative phase.

The equations above can be applied to the label propagation data of population HSC-2 and in principle allow for identification of all parameters plus the number of generations. Since the practical identifiability was missing, we included also the BrdU data and wrote analogous equation for the incorporation of BrdU label generation-wise. Such a model can only describe the data if the number of proliferation is high ($n > 11$), in which case the number of parameters also increases, having as a consequence the loss of information (Table 3.1). We conclude that the most simple explanation for the data is that proliferation is stochastic and cells do not remember how many generations they have gone through.

n	# data	# parameters	χ^2	$AICc$
stochastic events	51	18	45	103
1	51	19	113	176
2	51	20	95	163
3	51	21	78	151
4	51	22	68	148
5	51	23	60	147
6	51	24	54	148
7	51	25	51	153
8	51	26	50	160
9	51	27	49	170
10	51	28	46	176
11	51	29	46	190
12	51	30	42	195

Table 3.1: Ranking of models that differ in their topology.

Chapter 4

Comparing estimated parameters among different models

In Chapter 3 we modelled haematopoietic kinetics upon considering a refined definition for HSC and progenitor populations, according to which the populations we had introduced in Chapter 2 are heterogeneous. We nevertheless included the data for the homogeneous populations in the model of chapter 3 upon treating it as if the populations were heterogeneous. It is thus interesting to look in more detail into the interpretation of the results of different models when it comes to their performances with heterogeneous populations.

Let us consider, for example, a simple cascade where the stem cells A progressively differentiate into B, C and D. We compare such scheme with an hypothetical one where, for example, there is no experimental way to distinguish B and C, which are treated as an homogeneous population, BC, (Figure4.1). Next, we generate data with the heterogenous model. The type of data we want to fit together is the same as in Section 3.1, that is:

- BrdU data

- label propagation data

- size of compartments

For both models we compute the best fit and the confidence intervals on the parameters. Since the best fit with the homogeneous model reproduces the data, we do not need to suspect that this latter model is incorrect. Now, if we compared parameters, we would expect that parameters for population A are similar in both models, since it is the same population and we use exactly the same data. But this is not necessarily, the case, as shown in the
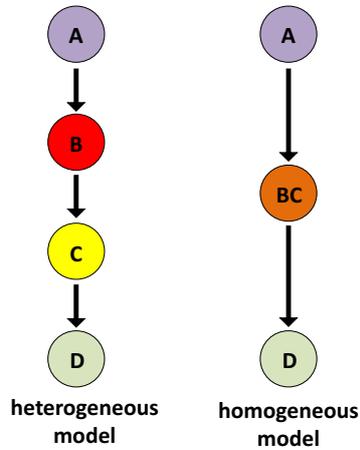
Figure 4.1: Toy model to compare heterogeneous populations. BC encompasses B and C together.

profiles obtained for the two models in Figure4.5:
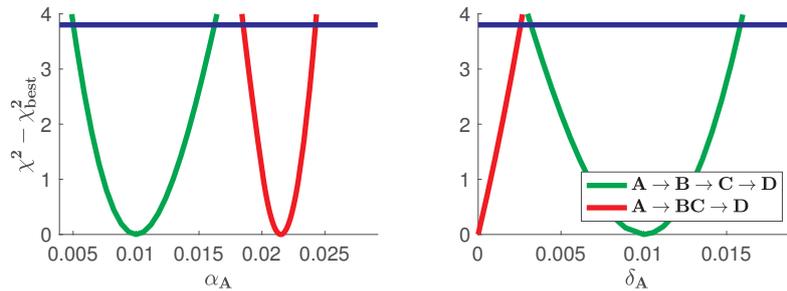
Figure 4.2: Comparing parameters profile for the heterogeneous (green) and the homogeneous (red) models. Data were simulated using the heterogeneous model

In figure 4.5 the differentiation rate estimated with the homogeneous model is higher than the actual differentiation rate, whereas the death rate estimated with the homogeneous is lower. This result is not surprising, as the differentiation rate of A is the rate at which cells of type B are produced from A. However, in the homogeneous model, B is lumped with a kinetically different compartment, giving the (wrong) impression that A feeds a larger compartment directly than it actually does.

Let us now ask the opposite question. What would happen if we modelled a homogeneous population as heterogeneous? For example, if BC were a homogeneous population, one could model BC as a cascade from B to C

having, for example, B very small compared to C, the flux from B to C very high compared to the influx in B, and C similar to BC in its kinetic rates. Thus if a model yields a very small population with fast efflux one can suspect it to be not sufficiently distinct kinetically from its progeny. Problems may arise when one measures data for the B and C populations and tries to model them as distinct and consecutive compartments.

In the following example, we generated data with the A→ BC → D model and fit it with the A→ B → C → D model, additionally fixing the ratio of C to B. The fit is very good, but parameter estimation depends on the assumed model (Figure 4.3):
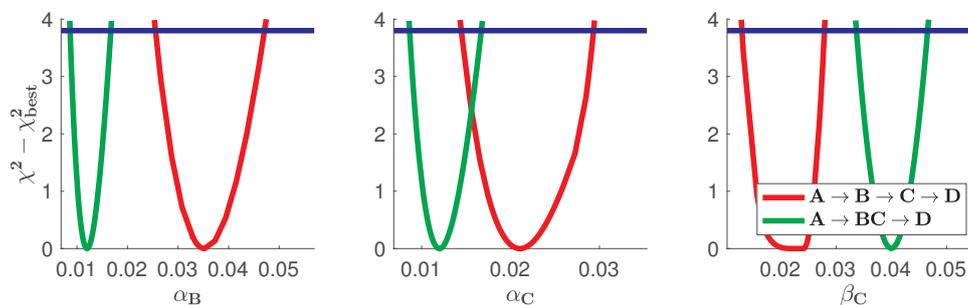
Figure 4.3: Comparing parameters profile for the heterogeneous (green) and the homogeneous (red) models. Data were simulated using the homogeneous model.

The differentiation rates from B and C are both overestimated, and in the case of B, the estimated rate is incompatible with the homogeneous model. Proliferation in C is also overestimated. In conclusion, the values of the parameters depend on the assumed compartment structure. However, we argue that combinations of parameters that represent physical quantities should show this dependance not.

We present in the following examples of quantities that we expect to be model-invariant. Having in mind what are the model invariant quantities turns out to be very useful when it comes to comparing the estimates of the same parameters performed with different experimental and/or theoretical setups.

4.1 Residence time

We introduced the concept of residence time earlier in Section 2.1.3, with respect to a single compartment. The residence time is the average time that a cell spends in a compartment before being lost by either death or

differentiation. We could analogously define the residence time of a cascade as the average time to exit the cascade for a progeny of a cell injected at the beginning of the cascade. Since this is the type of experiment we perform (injecting a labelled cell at the beginning of a differentiation cascade) and our model contains the information on the single residence times, it is intuitive to think that, if a model fits the data satisfactorily, then it should contain information on the residence time of the cascade regardless of the compartment chosen.

Let us consider a cascade of populations whose cell numbers are $N_i(t)$ and define as $q_i(t - t')$ the flux generated by a cell entering compartment i at t' and exiting compartment i at time t. Then the total flux out of compartment i would be:

$$j_i(t) = \int_0^t q_i(t - t')j_{i-1}(t')dt' \tag{4.1}$$

$$j_1(t) = \alpha_1 N_1(t) \tag{4.2}$$

We then Laplace transform $j_i(t)$ as:

$$j_i(s) = \int_0^\infty j_i(t)e^{-st} \ dt = q_i(s) * j_{i-1}(s) \tag{4.3}$$

$$j_1(s) = \frac{\alpha_1}{\lambda_1 - \alpha_1 - s} \tag{4.4}$$

thus $j_i(s) = \prod_{j=1}^i q_j(s)$. The residence time in a cascade starting with population 1 and ending with population M is:

$$\tau_{M|1} = \frac{\int_0^\infty t \ j_M(t) \ dt}{\int_0^\infty j_i(t) \ dt} = \frac{\dot{j}_M(s)|_{s=0}}{j_M(s = 0)} = \tag{4.5}$$

$$= \frac{\sum_{i=1}^M \dot{q}_i(s = 0) \prod_{j \neq i} q_i(s = 0)}{\prod_{i=1}^M q_i(s = 0)} = \sum_{i=1}^M \frac{\dot{q}_i(s = 0)}{q_i(s = 0)} = \tag{4.6}$$

$$= \sum_{i=1}^M \tau_i \tag{4.7}$$

In other words, the residence time of a cascade is the sum of the residence times of each population. We now compare the residence time out of ST-HSC for the homogeneous and heterogeneous models. The fact that hey match very well (Figure 4.4), suggest that our homogeneous and heterogeneous models are both good descriptions of the system.
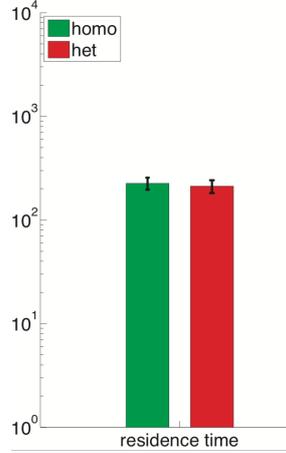
Figure 4.4: Comparing estimates and confidence bounds on the residence time in the ST-HSC population for the heterogeneous and the homogeneous models.

4.2 Number of proliferations and generations

Another interesting quantity is the number of generations a cell goes through before exiting the cascade. As for the residence time, this quantity is finite only if the residence time in each compartment of the cascade is finite.

We start calculating the number of divisions of a cell and its progeny in a compartment. It is clear that, if we knew how many cells exit the compartment for each injected cells, say N_{exit}, then we would also know the number of divisions, $N_{\text{prol}} = N_{1|1}^{\text{exit}} - 1$. For a population with kinetics parameters $\alpha_1, \lambda_1, \delta_1$ we have:

$$\dot{N_1}(t) = -\kappa_1 N_1(t) \tag{4.8}$$

$$N_1(0) = 1 \tag{4.9}$$

$$N_{\text{exit}} = \int_0^\infty (\alpha_1 + \delta_1) \ N_1(t) \ dt = \frac{\alpha_1 + \delta_1}{\kappa_1} \tag{4.10}$$

where $\dfrac{\alpha_1}{\kappa_1}$ are the cells lost by differentiation and $\dfrac{\delta_1}{\kappa_1}$ those lost by death.

If we now inject a cell in population 1 and look at the transition 1 to 2, we find:

$$N_{1|2}^{\text{exit}} = \frac{\delta_1}{\kappa_1} + \frac{\alpha_1}{\kappa_1} \frac{\alpha_2 + \delta_2}{\kappa_2} \tag{4.11}$$

The first addend is the number of cells lost by death in compartment 1, while the second is the loss from population 2 amplified by the flux amplification

50

from 1. In general, we have:

$$N_{1|M}^{\text{exit}} = \frac{\delta_1}{\kappa_1} + \sum_{i=2}^{M} \prod_{j=1}^{i-1} \frac{\alpha_j}{\kappa_j} \frac{\delta_i}{\kappa_i} + \prod_{j=1}^{M} \frac{\alpha_j}{\kappa_j} \tag{4.12}$$

We then compare the estimated number of divisions for the ST-HSC population, (Figure 4.5), which yields good agreement between homogeneous and heterogeneous models. Moreover, the estimated total flux of cells into ST-HSC, unlike the differentiation rate, is also expected to be invariant:
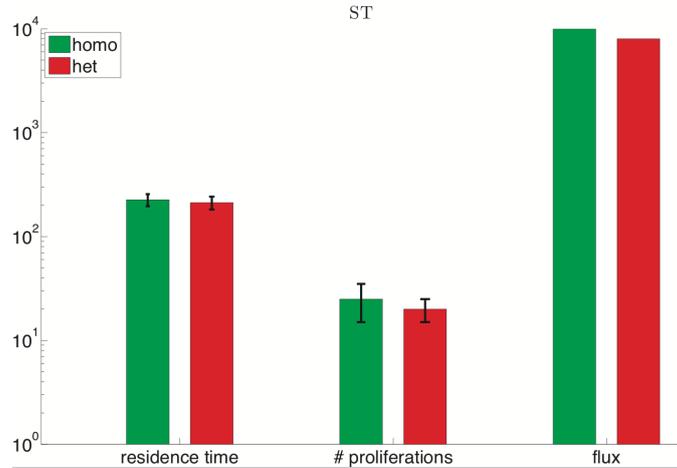
Figure 4.5: Comparing estimates and confidence bounds on the residence time, number of proliferations, differentiating flux of the ST-HSC population for the heterogeneous and the homogeneous models.

Again, lumping compartments is still a good description.

Finally, one could also raise the question: how many generations can a cell divide for before being depleted? The answer is not inferable directly from the above estimated number of divisions [Kay, 1965]. Let us consider for example a cell that undergoes 7 divisions, as in Figure 4.6. The number of generations can vary between a minimum of 3 (right panel) and a maximum of 7 (left panel). In order to compute the number of generations, we thus used a Gillespie algorithm that simulates the fate of a cell with the kinetic properties of ST-HSC. We found that the average maximum number of generations is 4 as calculated with the heterogeneous model and 5 calculated with the heterogeneous model. In conclusion, our data also contain the model-invariant information on the generations a cell undergoes in the ST-HSC population.
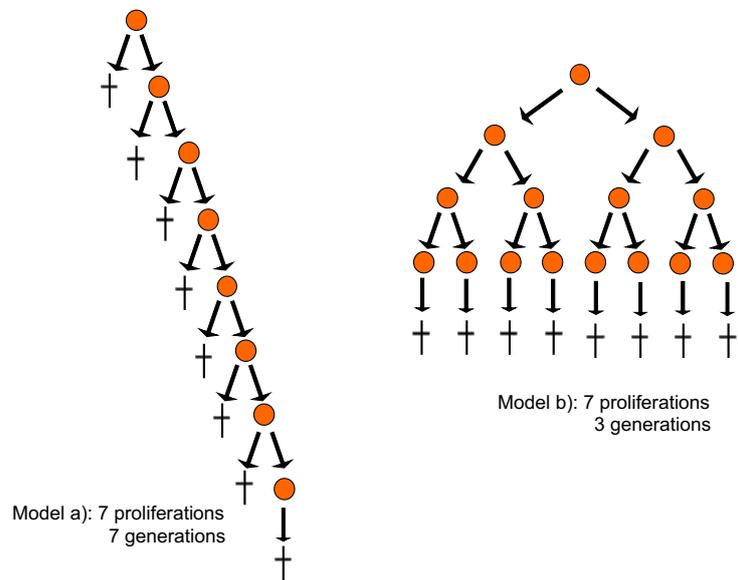
Figure 4.6: Different number of generations can be achieved with the same number of proliferations

Chapter 5

Non-stationary haematopoiesis after irradiation

In the previous chapters we built a model for stationary haematopoiesis and showed how the time course data on label progression and the ratio of populations sizes allow to estimate residence times and differentiation rates. We now wonder whether non-stationary data could be described by a similar approach, that is, a model with constant kinetic parameters and no explicit feedback mechanism. The only difference with the stationary model would be that the compartment sizes are allowed to vary over time The calculations performed in this chapter were in collaboration with master student Laura Obenauer.

For most perturbations this assumption might be too simple: for example, a sudden myeloablation would cause a feedback response on the intermediate progenitors of myeloid cells (hypothesis verified by data measured by PhD student Ann-Kathrin Schuon, personal communication). However, a strong perturbation that nearly depletes the mouse's blood system might be followed by a long period where stem cells reconstitute the blood at high activity, and it is not unlikely that such activity can be considered constant over extended time periods.

Ann-Kathrin Schuon measured time course data of population dynamics and label propagation in the *Tie2*-Cre mice [Busch et al., 2015] for up 16 weeks after sublethal irradiation. No deceleration in the growth of HSCs (a possible sign fir lack of feedback), while other progenitors did not recover their stationary size (as in [Li and Slayton, 2013]), see Figure 5.1. Our population dynamics model leads to an ODE system for the ratios of progeny-progenitor sizes $R_i(t)$ and the frequencies of labelled cells $f_i(t)$, as
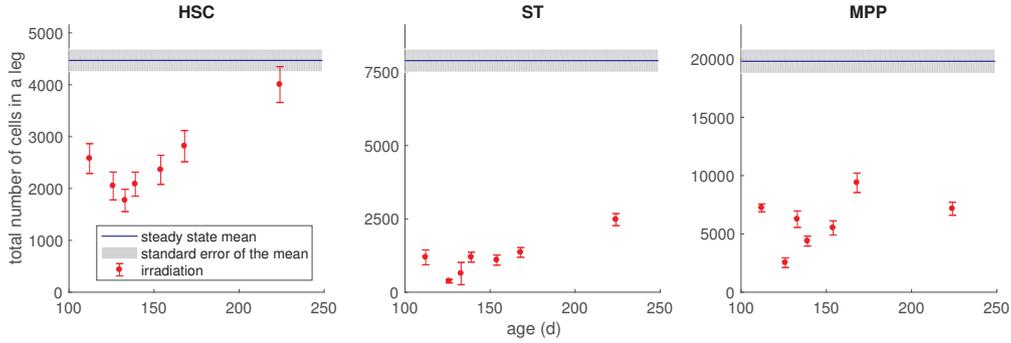
Figure 5.1: Population growth after irradiation. Data measured by Ann-Kathrin Schuon.

follows from Equations 2.11 and 2.16:

$$\dot{R}_i(t) = \alpha_{i-1} - \kappa_i \ R_i(t) - \alpha_{i-2} \frac{R_i(t)}{R_{i-1}(t)} + \kappa_{i-1} \ R_i(t) \tag{5.1}$$

$$\dot{f}_i(t) = \frac{\alpha_{i-1}}{R_i(t)} \ (f_{i-1}(t) - f_i(t)) \tag{5.2}$$

It is worth noting that for non-steady state we obtain a closed ODE system for the derived variables $R_i(t)$ and $f_i(t)$, with the primary variables being cell numbers. This system reduces to the steady state model for $\dot{R}_i(t) = 0 \ \forall i$. The identifiable parameters are, once again, the residence times and the differentiation rates.

In Figure 5.2 we show the normalised labelling frequency in ST-HSC as compared to the model prediction in the case of steady state kinetics. Since this latter underestimates the actual data, one might conclude that the flux from stem cells is increased.

Let us compare the differential equation for $f_1(t)$, the first progeny of a stem cells population labelled with 0, and its steady state counterpart, $f_1^{\mathrm{ss}}(t)$:

$$\dot{f}_1(t) = \frac{\alpha_0}{\dfrac{\alpha_0}{\kappa_1 - \kappa_0} + \left(R_1(0) - \dfrac{\alpha_0}{\kappa_1 - \kappa_0}\right) \ e^{(\kappa_1 - \kappa_0) \ t}} \ (1 - f_1(t)) \tag{5.3}$$

$$\dot{f}_1^{\mathrm{ss}}(t) = \frac{\alpha_0^{\mathrm{ss}}}{\dfrac{\alpha_0^{\mathrm{ss}}}{\kappa_1^{\mathrm{ss}}} + \left(R_1(0) - \dfrac{\alpha_0^{\mathrm{ss}}}{\kappa_1^{\mathrm{ss}}}\right) \ e^{k_1^{\mathrm{ss}} \ t}} \ (1 - f_1^{\mathrm{ss}}(t)) \tag{5.4}$$

where we have plugged into Equation 5.2 the analytical solution of Equation 5.1. We see that $\alpha_0 \gtreqless \alpha_0^{\mathrm{ss}}$ is a sufficient condition for $f_1(0) \gtreqless f_1^{\mathrm{ss}}(0)$, but the
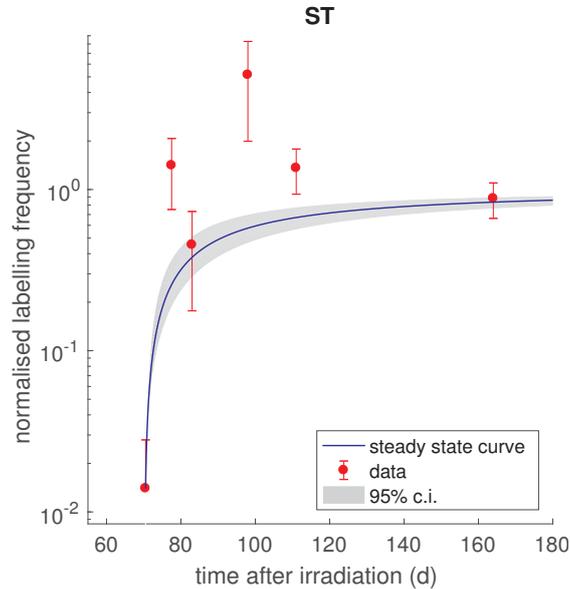
Figure 5.2: Labelling propagation in the ST-HSC population. Comparison between steady state prediction and measured data.

inequality can change over time[1]. Moreover, the condition is not necessary, as a change in the kinetics of the stem cells' progeny could also determine whether the labelling frequency is higher or lower than the steady state labelling frequency. We expect that fitting the model to the data will uncover which parameters actually change after challenge.

As we did for the steady state case, we implemented also cell-cycle-related data in order to estimate the proliferation rate. The data available in this case was measured by Ann-Kathrin Schuon and consists of the time course accumulation and depletion of EdU, a thymidine analogue similar to BrdU. EdU was administrated in one intraperitonal injection to tamoxifen-treated *Tie2*-Cre mice, and the EdU$^+$, *Tie2*$^{+/-}$ cells were measured over time for up to one week and for each phase of the cell cycle (Figure 5.3).

Unlike in Chapter 3, we are now facing a scenario where:

- The available data for the accumulation of EdU were measured over a time span of one day. In this period the fraction of EdU-positive cells increases fast due to the fast incorporation of the molecules into the

---

[1]In Section 2 we saw the simpler case of labelling frequencies kinetics in ageing mice. In that case $R_i(0) = \dfrac{\alpha^{\text{ss}}}{\kappa_0^{\text{ss}}}$, thus the equations simplify and show that the decrease of either HSC or ST-HSC output is a necessary and sufficient condition to explain the data.
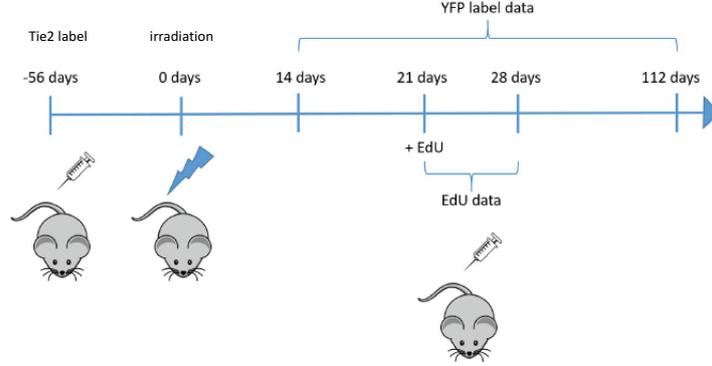
Figure 5.3: Scheme of experiment combining the *Tie2*-Cre induction, irradiation and EdU administration.

DNA of cells in S phase. We cannot thus, as we did before, assume that cells get positive if and only if they go through one division, and thus we need to get back to a cell cycle phase-resolved model.

- From day one to seven days EdU is clearly used up and eliminated, so that the fraction of positive cells decreases due to division. This must be incorporated in the model, too.

For the uptake part we used the scheme explained in Chapter 3 for the HSCs. As discussed in Section 3.2, we write the equations:

$$\dot{N}^+_{\mathrm{G_0/G_1}}(t) = -(a + d1 + m1)\ N^+_{\mathrm{G_0/G_1}}(t) + 2\ c\ N^+_{\mathrm{G_2/M}}(t) \tag{5.5}$$

$$\dot{N}^+_{\mathrm{S_1}}(t) = a\ N^+_{\mathrm{G_0/G_1}}(t) - b\ N^+_{\mathrm{S_1}}(t) \tag{5.6}$$

$$\dot{N}^+_{\mathrm{S_i}}(t) = b\ (N_{\mathrm{S^+_{i-1}}}(t) + N_{\mathrm{S^-_{i-1}}}(t) - N_{\mathrm{S^+_i}}(t)) \tag{5.7}$$

$$\dot{N}^+_{\mathrm{G_2/M}}(t) = -(c + d2 + m2)\ N^+_{\mathrm{G_2/M}}(t) + b\ (\mathrm{S}^+_n(t) + \mathrm{S}^-_n(t)) \tag{5.8}$$

where a is the rate of progression from $\mathrm{G_1/G_0}$ to S, b is the rate at which a cell synthesises enough DNA to became EdU positive, c is the rate at which a cell exit mitosis to form two daughter cells in $\mathrm{G_2/M}$ phase, $d_{1,2}$ and $m_{1,2}$ are the rates at which a differentiation or a death event take place in the $\mathrm{G_1/G_0}$ or $\mathrm{G_2/M}$ phases respectively. In order to compare these rates with the way we have previously defined the proliferation rate we need to consider

the phases distribution according to:

$$\lambda_{\text{HSC}} = c \ \frac{N_{\text{G}_2/\text{M}}(t)}{N_{\text{G}_0/\text{G}_1}(t) + N_{\text{S}}(t) + N_{\text{G}_2/\text{M}}(t)} = \ const \qquad (5.9)$$

$$\alpha_{\text{HSC}} = \frac{d_1 \ N_{\text{G}_0/\text{G}_1}(t) + d_2 \ N_{\text{G}_2/\text{M}}(t)}{N_{\text{G}_0/\text{G}_1}(t) + N_{\text{S}}(t) + N_{\text{G}_2/\text{M}}(t)} = \ const \qquad (5.10)$$

$$\delta_{\text{HSC}} = \frac{m_1 \ N_{\text{G}_0/\text{G}_1}(t) + m_2 \ N_{\text{G}_2/M}(t)}{N_{\text{G}_0/\text{G}_1}(t) + N_{\text{S}}(t) + N_{\text{G}_2/\text{M}}(t)} = \ const \qquad (5.11)$$

The rates have been set to be constant as assumed earlier. This implies that the fractions of cells in each phase are constant over time, or, in other words, that HSC is growing at constant rate (specifically, we set the time zero for modelling at 14 days after irradiation, corresponding to the first data point, so we can assume that the transient phase is already over).

For the dilution part we reasoned as follows: each time a cell divides its labelled DNA decreases by one half. So a cell becomes EdU negative when the fraction of labelled DNA gets below $\frac{1}{n}$, the minimal fraction to be considered as positive. We estimated the quantity $Q(t)$ of accumulated positive DNA during the uptake as follows:

$$\dot{Q}(t) = b \ S(t) \ \frac{1 - Q(t)}{2} \qquad (5.12)$$

and then calculated the minimum number of divisions to become negative as $m \geq \log_2(n \ Q_{\text{dil}})$, where Q_{dil} is the quantity of positive DNA when dilution starts. n is a parameter of the model. The dilution equations read:

$$\dot{N}^+_{\text{G}_0/\text{G}_1,j}(t) = -(a + d1 + m1) \ N^+_{\text{G}_0/\text{G}_1,j}(t) + 2 \ c \ N^+_{\text{G}_2/\text{M},j-1}(t) \qquad (5.13)$$

$$\dot{N}^+_{\text{S}_1,j}(t) = a \ N^+_{\text{G}_0/\text{G}_1,j}(t) - b \ N^+_{\text{S}_1,j}(t) \qquad (5.14)$$

$$\dot{N}^+_{\text{S}_i,j}(t) = b \ (N^+_{\text{S}_{i-1},j}(t) + N^-_{\text{S}_{i-1},j}(t) - N^+_{\text{S}_i,j}(t)) \qquad (5.15)$$

$$\dot{N}^+_{\text{G}_2/\text{M},j}(t) = -(c + d2 + m2) \ N^+_{\text{G}_2/\text{M},j}(t) + b \ (\text{S}^+_{n,j}(t) + \text{S}^-_{n,j}(t)) \qquad (5.16)$$

where $j \in 0 - n$ is the number of divisions a cell has gone through after exhaustion of EdU (exhaustion time to be estimated in the model). All mentioned parameters are structurally identifiable with our data set, which includes: fraction of EdU positive cells over time, fraction of cells that are in S phase and that are EdU positive in S phase during uptake. The model fit to the data is shown in Figure 5.4, and the parameters value in Table 5.

From this parameter analysis, we conclude that irradiation speeds up the kinetic parameters in the HSC: proliferation is accelerated by an order of magnitude and differentiation by a factor 2. Confidence bounds on the death
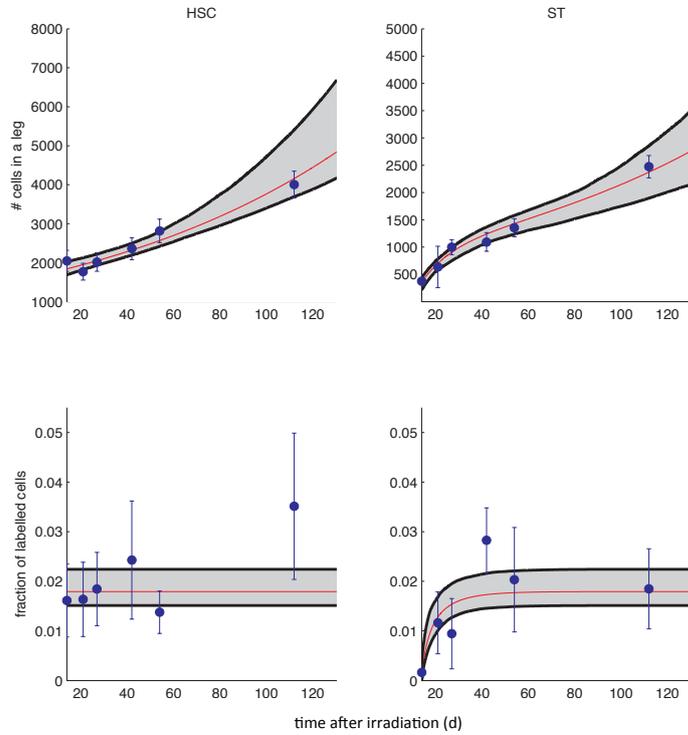
Figure 5.4: Fit to the population growth and label propagation after irradiation. Blu: data, red: model, grey: 95% bootstrap confidence interval on the model. Data measured by Ann-Kathrin Schuon.
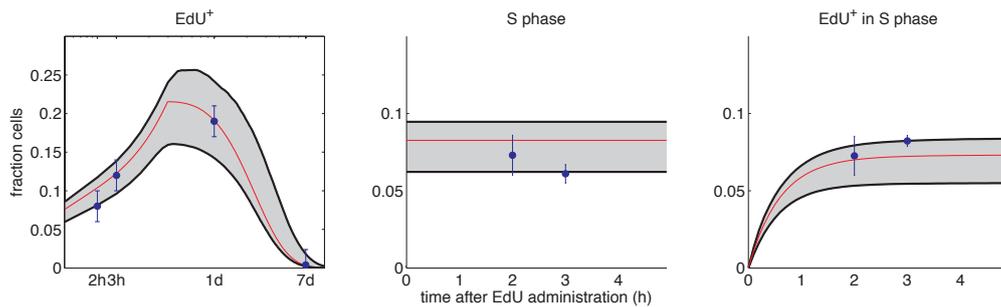
Figure 5.5: Fit to the EdU time course data and phases distribution over time. Blu: data, red: model, grey: 95% bootstrap confidence interval on the model. Data measured by Ann-Kathrin Schuon, model in collaboration with Laura Obenauer.

58

	steady-state	post-irradiation
λ_{HSC}	0.036 [0.028 0.045] d^{-1}	0.24 [0.19 0.28]d^{-1}
α_{HSC}	0.008 [0.006 0.013] d^{-1}	0.05 [0.02 0.10]d^{-1}
δ_{HSC}	0.03 [0.02 0.04] d^{-1}	0.14 [0.02 0.19]d^{-1}
κ_{ST}	0.003 [0.002 0.004] d^{-1}	0.07 [0.05 0.14]d^{-1}

Table 5.1: Comparison between steady state and post-irradiation kinetics.

rate do not allow to state whether there is a change in comparison to steady state. Residence time in ST-HSC drops by an order of magnitude, suggesting that this compartment is now rather transient. In conclusion, after a strong perturbation such as irradiation, the most immature stem cells are massively recruited to reconstitute the blood system upon accelerating their kinetics.

5.1 Refining the dilution model

In the previous section we modelled the EdU dilution upon assuming that at each round of division the quantity of labelled DNA is reduced by a factor one half, and thus after a certain number of divisions a cell will be negative for EdU in the FACS. This is not rigorous, if one considers that DNA is organised in chromosomes and that chromosomes are randomly segregated.

To illustrate the problem let us consider the hypothetical case of cells with two chromosomes only, labelled with A and B (Figure 5.6). Let us assume that at the beginning DNA is maximally (100%) labelled, (Figure 5.6a), and that we need 25% of the maximum to consider a cell labelled. For the moment we will ignore differentiation and death. After one division there is only one possibility: the two daughter cells have each 50% of labelled DNA and thus appear both EdU positive (Figure 5.6a). Now consider one such cell in generation 1, and refer to the two strands in each chromosome as 1 and 2. Upon replicating its DNA, (Figure 5.6c), the total content of DNA will drop to 25%, but the number of positive cells can be either one or two depending on how the chromosomes are segregated, (Figure 5.6d). So, after two cycles we already have negative cells, and after three or more divisions we still have positive cells. If one does not consider the chromosomal organisation, as we did in the previous section, there would be still 100% of positive cells after 2 divisions and no positive cells after three divisions, which a very different outcome. However, this discrepancy gets smaller if the fraction of labelled DNA needed to be detected as positive is higher and so is the number of chromosomes.
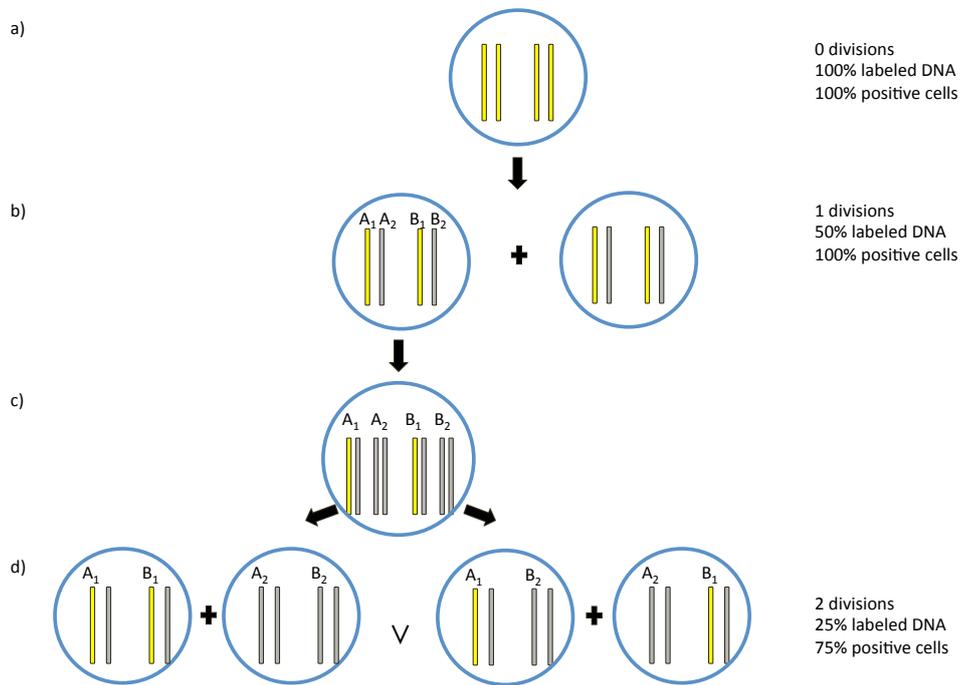
Figure 5.6: Label dilution upon cell cycle.

In the real case of a mouse, there are 40 chromosomes and one can estimate upon modelling, as in the previous section, that the threshold for positivity is 12.5%, corresponding to 10 labelled strands. We compared the chromosome-resolved and coarse-grained models for the dilution of EdU in HSCs, and found little discrepancy in the estimated cell cycle time. We thus keep the coarse-grained model since it has fewer equations (4 instead of 42).

Chapter 6

Discussion

In this thesis we presented an application of a deterministic mathematical framework to infer quantitative information on normal and challenged haematopoiesis from stem cells *in vivo* from fate mapping and cell cycle time course data.

The steady state role of HSCs. We addressed this point in three parts. First we asked whether HSCs are the major sustainers of haematopoiesis. We fit a population-based model to data representing the flow of a permanent genetic label induced in the HSC population and propagated in time into the HSC's direct progeny, ST-HSC [Busch et al., 2015]. The information extrapolated from this data is the residence time of a cell in population ST-HSC, that is, the time it takes for a cell of type ST-HSC to be lost by either death or differentiation. Since we found this residence time to be high compared to the lifespan of a mouse ($\sim 1y$), we concluded that ST-HSCs are the major source of normal haematopoiesis. One can better grasp this property by imagining that a certain population is all at once depleted and no feedback mechanism is triggered. The residence time is the expected time before the number of cells downstream of the ablated population noticeably drops. This would be, for example, around one year if HSCs were depleted or two months if ST-HSCs were. This finding is coherent with [Schoedel et al., 2016], who measured no significant changes in stationary haematopoiesis if HSCs are depleted to non-measurable exstent. However, our results contradict the faster equilibration kinetics of both ST-HSC and MPP inferred by [Sawai et al., 2016] with a theoretical and experimental framework similar to ours, which would rather suggest the prominent role of HSCs in the maintenance of haematopoiesis. However, [Sawai et al., 2016] performed bone marrow biopsy, which could activate stem cells and progenitots.

We then quantified the contribution in terms of frequency of differentia-

tion of the HSCs to normal haematopoiesis. We relied again on the fate mapping data, using as additional information the stationary ratios of total cell numbers in each population, easily included in the steady state model. We found a very infrequent recruitment of HSCs into differentiation (of the order of one differentiation per one hundred days per cell). Nevertheless, this contribution is significantly non zero, implying that HSCs still fuel haematopoiesis with a steady although rare contribution. Similar estimates were obtained by [Mackey, 2001] with a deterministic model of BrdU label propagation data measured by [Bradford, 1997] and [Cheshier et al., 1999]. Also [Sun et al., 2014] qualitatively concluded that stem cells only rarely contribute to haematopoiesis, using barcoding and fate mapping of HSCs and progeny.

Finally, we investigated the number of HSCs that are active in the observation period, meaning the number of cells that differentiated at least once, and thus actively contributed to haematopoiesis, in the time span of the experiment. To this end, we statistically analysed the results of a limited dilution experiment, where few HSCs are labelled and the frequency of mice with no labelled progeny is recovered over time. As no such mouse was found, no matter the time point and the dilution, we concluded that potentially all cells are active, but at least 30% must be.

This result is clearly antithetic to the outcome of several studies based on repopulation of transplanted stem cells after lethal irradiation [Gerrits et al., 2010] [Lu et al., 2011], according to which less than one hundred HSCs contribute to haematopoiesis (less than 0.1% of the steady state number of HSCs), suggesting that the steady state scenario is very different. One obvious difference is the number of engrafted cells (statistically 1/3), but still the majority of cells seems to be inactive post-transplantation, which would hint at a competition for resources and access to environmental cues. Our finding is again coherent with [Sun et al., 2014], according to whom physiological haematopoiesis is a polyclonal process.

We additionally computed via mathematical modelling of BrdU accumulation that at least 20% of HSCs must be actively cycling. In order to compare this result with publications which estimated the fraction of cells going through the cell cycle over long enough observation periods, we first need to clarify our concept of quiescence. Our idea is that an active cell can both differentiate and proliferate, while a quiescent cell is not recruited into haematopoiesis at all. In the literature, quiescent cells are usually defined as cells that are in the G_0 phase of the cell cycle. In our understanding, a cell in G_0 can be stochastically recruited into the cell cycle also in normal conditions, although this eventuality can be so unlikely that it does not take place along the duration of an experiment. Thus there is no contradiction

between our outcome on many (potentially all) HSCs being active over one year and the measured fraction of cells being in G_0, estimated to be over 90% [Oguro et alii, 2013]. Being in G_0 is also not in contradiction with models that assume differentiation taking place in G_0 as in [Mackey, 2001]. Comparing our result with literature, [Cheshier et al., 1999] administrated BrdU 6 months and eventually all cells were labelled. Similarly, in [Bernitz et al., 2016] a 2 years long doxocyclin chase was performed with a H2B-GFP mouse model, together with comparison of peak fluorescence of the GFP, and only a small fraction of HSCs was estimated to have not proliferated at all, and around 3% of HSCs have divided up to a maximum of four times. On the other hand, [Takizawa et al., 2011] determined with both BrdU and CFSE a small fraction (1%) of HSCs that do not divide at all, although they transplanted cells in a non irradiated mouse and did not monitor it for longer than 20 weeks. From these data, it is likely that over long time periods nearly all HSC will go through the cell cycle. IN addition, we have found that many HSCs differentiation oves long time periods. Taken together, these data indicate that the HSCs are both dividing and differentiating but do so with low rates per cell. It is unclear whether an individual HSC divides and proliferates, but the observation that the fate mapping label is maintained in HSCs and propagated eventually to all mature cells proliferation suggests so.

Activity of progeny With a similar framework we inferred residence times and differentiation frequency for several compartments downstream to HSC and ST-HSC. Residence time turned out to be smaller in the myeloid populations, who thus work rather as transient amplifiers of flux exception than as reservoirs for generation of haematopoietic output. On the other end, lymphoid progenitors are rather long lived, which is in contrast to previously reported results, that state that the thymus can only self-renew in the absence of input [Martins et al., 2012]. The production of a great numbers of mature blood cells, especially monocytes and erythrocytes, is achieved thanks to large amplification via proliferation and differentiation, whose rates increase with lineage specification, particularly in the myeloid branch. Very few parameters have been previously estimated for the differentiation frequency of downstream population. Our results on T and B cells are compatible with such results (reported in [Manesso et al., 2013]).

Compartmental model, topology and branching. In order to infer parameters from fate mapping we developed a model based on the idea of haematopoiesis as a flow of cells through separate compartments characterised by different kinetic parameters, and connected according to the classical view of haematopoiesis [Morrison and Weissman, 1994] [Akashi et al., 2000] [Kondo et al., 1997] [Pei et al., 2017]. This approach proved to be able to describe the data and thus we found no strong contradiction to it. however, this assumption strongly influences the conclusions we drew.

63

On the one end, the definitions of compartments is based on surface marker expression and not on molecular regulation, which is more likely to correlate to proliferation and differentiation potential. In [Laurenti and Dick, 2012], for example, bioinformatic analysis of transcriptional dynamics suggests that there is no such rigid distinction among haematopoietic compartments. This scenario would be rather modelled by a continuum of microstates [Karamitros et al., 2018], groupable in macrostates to resemble the more familiar idea of compartment, as in [Stumpf et al., 2017] .

The most influential assumption is, though, the lineage topology. After having explored the implications of the classical model we tried to infer information on what should be the most likely topology. This should in principle be hard with population average data, but the fate mapping data contains information on at least the progenitor-progeny relationship, meaning that one can say whether a population is likely to be downstream to another upon considering that a downstream population should not have a higher labelling frequency than any of its progenitors. Of course, all estimated rates are representative of an average behaviour and they do not account for what a single cell actually does. We compared different possible topologies, keeping the usual assumption that no back differentiation takes place and limiting ourselves to the myeloid lineage. Based on statistics and information theory, we found that additional differentiation pathways from the stem cells, and at least from ST-HSCs, to all myeloid progenitors and differentiated cells describe the data better than the classical linear model, with its progressive loss of multipotency. This refined lineage topology is consistent with the conclusions of several papers postulating the existence of alternative pathways originating from either stem cells or their immediate progenitors [Perié et al., 2014] [Yamamoto et al., 2013] and [Perié et al., 2015], and particularly the existence of an erythroid bias [Nutt et al., 2005] [Lai and Kondo, 2006] [Guo et al., 2013] [Notta et al., 2016] [Hoppe et al., 2016]. On the other hand, these conclusions rely always on perturbed conditions (apart from megakaryocyte-biased stem cells, that has been recently described for native conditions [Rodriguez-Fraticelli et al., 2018]). One could speculate that the existence of the alternative pathways is more evident in non-stationary challenged situations, where blood production is urgently needed and additional differentiation pathways may be enhanced, though they are likely to exist also in steady state, as we were able to confirm with our framework. Note that a different topology implies a different estimation for the residence times, yielding longer residence times for myeloid progenitors, which in this view would rather serve as reservoirs than as amplifiers for blood production.

Another type of topology we considered was to take the linear model as true but anticipating the branching at the level of ST-HSC, instead of MPP,

as suggested by [Oguro et al., 2013]. Here we found the interesting result that the differentiation rate towards CMP is no longer enormous compared to CLP, although the total number of cells differentiating to become CMP per unit time is the same. So, the bias towards the myeloid production is no longer a matter of cells being primed towards one lineage or the other, but rather the strong amplifications is guaranteed by the transient amplifying compartments (i.e., selective proliferation of myeloid progenitors).

Mathematical model of BrdU uptake. Several mathematical models have been implemented to model a stem cell population taking up a cell-cycle-related label. When the steady state for the number of cells can be assumed and BrdU is continually fed to the mice, one can simply assume that, as a cell divides, it becomes BrdU positive [Mohri et al., 1998] [Ganusov and Boer, 2012], and stays positive because BrdU is never diluted. This simple model has three potential problems:

- The model assumes that having divided at least once is a necessary and sufficient condition for a cell to become BrdU positive. In reality a necessary condition should be going through the S phase of the cell cycle, where the DNA replication and thus the label incorporation takes place. The discrepancy to the usual model is only relevant if the measurement time is short compared to the length of S, G_2 and M phases, in which case one would measure a non zero percentage of positive cells (corresponding to the fraction of cells in S phase) before any division takes place, and cells that are in G_2 or M phases divide without having become positive. In a setup where most of the cell cycle time is spent in G_1 any cell replicating its DNA is also dividing before the first time point is measured and the number of cells in G_2 or M phases is negligible. This is the case for almost all populations of interest for us, with the exception of HPC-1, as we have shown via a priori and a posteriori simulations. Finally, is one replication round a sufficient condition for being detected as positive? With one division, at most 50% of the DNA is labelled, though there is some evidence for BrdU incorporation being stochastic in haematopoietic cells including stem cells [Takizawa et al., 2011]. This problem has been deeply analysed in [Schittler et al., 2013]. Although stochasticity would make a difference at the level of the single cell, one could reason that it does not at the population level, provided that on average the uptake of label after the first division is above the threshold of detectability. We do not have an estimate for such threshold, but we can estimate what this should be at least. According to several studies ([Foudi et al., 2008] and [Wilson et al., 2008] for example), it takes about 5 divisions to dilute the label below the threshold of detectability. A positive cell has, of course, maximum 100% of labelled

DNA. After 4 divisions this would reduce to maximum 6% (cell is still positive) [van der Wath RC, 2009]. Thus a cell should only label 6% out of a maximum of 50% in order to become positive, which supports the idea that a single round of replication renders a cell BrdU positive.

- Another possible problem would arise if not all HSCs were be active. We discussed in the previous section why we believe that many HSCs are active. Also, upon introducing a quiescent population within HSCs in our model, we could determine a high lower bound for the number of contributing stem cells (95% for HSC-1 and 40% for HSC-2). In other words, having too many quiescent cells is not compatible with the data.

- It has been pointed out that BrdU can have a mitogenic effect on HSCs, that is BrdU accelerates the rate at which cells proliferate ([Wilson et al., 2008] and [Takizawa et al., 2011]). This does not seem to be the case for the data we worked on. In [Oguro et al., 2013], the authors use a similar experimental procedure and find similar proliferative kinetics as in [Kiel et al., 2007], where the authors checked whether the label incorporation due to the BrdU induced damage affects the fluorescence due to normal replication and found a negative result. This is stated more generally in [Spalding et al., 2005], where the authors proved that DNA repair does not add a detectable amount of incorporated fluorescent nucleotides.

Note also that we are only interested in the cells being ether positive or negative, and not in the intensity of the label (quantity of incorporated BrdU). This latter information is only useful if one either has measured the average intensity of the label over time [Bernitz et al., 2016] or is interested in modelling the dilution of the label over the chase period ([van der Wath RC, 2009]), in which case knowing the initial condition is crucial. So, we were able to infer proliferation frequency for the most immature haematopoietic populations. The trend we observed earlier with the net proliferation rates was confirmed, that is, proliferation increases with maturation. Our estimate for the stem cells is in the order of magnitude previously estimated [Mackey, 2001] [Wilson et al., 2008] [van der Wath RC, 2009] [Foudi et al., 2008].

Another interesting point is that previous cell cycle models have not simultaneously considered all possible fates a cell can undergo [van der Wath RC, 2009], [Wilson et al., 2008]: symmetric self-renewing division, asymmetric division, symmetric differentiating division, differentiation without division, death. It is easy to see that for population dynamics with constant rates such a detailed parametrisation is redundant, since every combination of fates can be reduced to a combination of a net rate of division term and a

term at which progeny is produced (although it does make a difference to have different time dependence in the rates in non stationary situation like in [Marciniak-Czochra et al., 2008] or when it comes to stochastic approaches [van der Wath RC, 2009], which we are not considering). Our result that symmetric self-renewing division must be the main mechanism of stem cell proliferation in the stem cells is in agreement with [Ito and Ito, 2016]. Having many symmetric divisions is an efficient way of expanding a cell pool in case of emergency, but could also lead to the expansion of a mutant clone as suggested in [Dingli et al., 2007a]. Also, it implies that the homeostasis is regulated rather at the population level than at the level of the single cell [MacArthur and Lemischka, 2013], and that proliferation is uncoupled from differentiation.

Invariants. We stressed in this work that fate mapping data contains information about the residence time of a cell in a population. On the other hand we presented two frameworks that describe the same set of data, and that consider the stem cells as homogeneous or heterogeneous populations respectively. We also explained how different assumed topologies can lead to estimation of different parameters. So the question naturally arises: if both models are good description of the physical system, should they predict the same residence time in a differentiating cascade?

In order to provide an answer I developed, together with Dr. Nils Becker (personal communication), a theory to calculate the residence time of a cell in a proliferating cascade. The result is the additivity of residence times of the single populations, and thus the independence of the total residence time from their order. I then compared the estimations performed on the data and found no contradiction in the result of the heterogeneous and homogeneous models.

Furthermore, I also developed a simple model to compute the number of proliferations of a cell and its progeny while flowing through a cascade, and again found no contradictions between the models' prediction for this quantity. This was also expected, since the number of proliferations is connected to the residence time and the proliferation rate, so to two informations contained in the data independently of the model.

Finally I simulated with a Gillespie algorithm that the average maximum number of generations reached by a cell in the cascade up to HPC-2 is about 25, in a similar range as in [Mackey, 2001] and [Dingli et al., 2007b].

Non steady state The model we developed for steady state is easily adjustable to describe perturbed scenarios, provided that the total number

67

of cells in each compartment of interest is also measured over time. Also, we integrated cell cycle information in the form of accumulation of EdU in each phase of the cell cycle, which were modelled considering in detailed the cell cycle progression. We applied this model to label progression and population dynamics in the most immature haematopoietic populations after sub-lethal irradiation and inferred a strong recruitment of HSCs for several weeks post irradiation, both as an acceleration of differentiation and as an increase in proliferative activity. Additionally, HSCs recovered completely but their progeny did not, even 16 weeks after irradiation as in [Li and Slayton, 2013], suggesting a major role for HSC in fueling haematopoiesis after irradiation. Finally, time spent in ST-HSC drops considerably, which suggests that ST-HSC is rather a transient compartment after irradiation, in contrast to their major reconstituting role in transplantation ([Yang et al., 2005]).

Parameter estimation. In this thesis we stressed tquantitative inference from experimental data. This means building models whose parameters can at least in principle (structurally) be identified from the available data [Buchholz et al., 2013] [Raue et al., 2009]. Others, for example [Mackey, 2001], have developed models with a redundancy of parameters, or have not estimated bounds on parameters [Sawai et al., 2016], or have inferred parameters based on an assumed behaviour the system should have in hypothetical conditions (mathematical stability [Marciniak-Czochra et al., 2008], efficient recovery after stress [Manesso et al., 2013]).
In our case, these are the learnable parameters (structural identifiability was checked with the Taylor expansion method [Pohjanpalo, 1978]):

- residence time, from steady state fate mapping time course data

- residence time, differentiation rate and net proliferation rate, from steady state fate mapping time course and compartment size ratios data

- residence time and proliferation rate , from steady state cell cycle time course and compartment size ratios data

- proliferation, differentiation and death rate from steady state fate mapping time course, compartment size ratios and cell cycle time course data

Using the data in the last point we were also able to practically identify almost all parameters, up to finding at least a meaningful lower bound. We could also draw conclusions on the mechanism of proliferation of the haematopoietic stem cells. In addition, we were able to simultaneously fit all the types of data, showing that they are compatible, although measurements

have been performed by different groups on different mice models ([Busch et al., 2015], [Oguro et al., 2013]).

Chapter 7

Methods

All fits were performed via the *lsqnonlin* inbuilt function of the MATLAB software, which uses the trust region reflective algorithm to find a local minimum of the cost function:

$$\chi^2(\mathbf{p}) = \sum_{i}^{n} \left(\frac{o_i - m_i(\mathbf{p})}{s_i} \right)^2 \tag{7.1}$$

from a given initial point in the space sample of the parameters $\mathbf{p}$. n is the number of data. o_i is the set of observables at the i-th time points, s_i the set of uncertainty at the i-th time points, s_i the set of the model predictions evaluated at the i-th time points and at the parameter vector $\mathbf{p}$.

The uncertainty is either the standard error on the mean if there are enough measurements per time point, the squared root of the pulled variance if data had to be pooled from different time points, or a gaussian noise on simulated data as in chapter 4.

If not differently stated, all confidence bounds on parameters were computed with the method of profile likelihood [Raue et al., 2009]. The profile of parameter i is the set of values of p_i for which the corresponding minimum minus the best fit minimum of the cost function is lower than 3.8 (5% quantile of the distribution of the reduced chi-squared). Each minimum is computed upon fixing the values of p_i and optimise on the remaining components of the parameter vector.

Confidence bounds on some quantities, for example the ratio of a population size with respect to HSC, were computed via bootstrap [B and (1993)., ages] since it was too complex to reparametrise the model to have them as an independent parameters.

Confidence bounds on models were also computed via non parametric bootstrap. Several thousands of iterations are performed extracting each time

a new data set from the measured one, and computing the correspondent vector of parameters that minimises the new cost function. Model curves are generated from each estimated parameters vector, and the 2.5% quantiles discarded per each time points.

The correct Akaike information criterion AICc [Burnham and Anderson., 2002] is used in Section 2.3 to compare different models for the same dataset, particularly to check whether complex models with more parameters add information or rather overfit the data. The formula for the AICc is:

$$AICc(\chi^2, n, k) = 2\ k + \chi^2 + 2\ k \frac{k+1}{n-k-1} \tag{7.2}$$

where k is the number of parameters. Models with the lowest index are more informative. Models with similar AICc are considered indistinguishable.

Gillespie algorithm [Gillespie, 1976] was used in Section 4 to compute the average number of generations a cells can reach at most given the proliferation and net efflux rates, and thus the number of proliferations of the cell and all its progeny before exiting the compartment. Since the number of combinatorial trees given the number of proliferations is a difficult quantity to compute, we just simulate thousands of different events starting from a cell and stored, and then averaged, the generation numbers.

We used simple running averages to smoothen noisy time course data and uncover their trend. Given a set of data $\{x_i\}$ corresponding to the set of time points $\{t_i\}$ and an arbitrary number n of, for every n consecutive time points the average m_i of the corresponding data is computed:

$$m_i = \frac{1}{n} \sum_{j=1+m}^{n+m} x_j \tag{7.3}$$

The moving averages is the set $\{m_i\}$. n is chosen arbitrarily to have smooth curves. In our case it was picked $n = 40$.

Identifiability analysis is based on the Taylor series approach for linear systems. Details are discussed in paragraph 3.2.

Chapter 8

Supplementary Material

2.1.3

We do not have time course data of the number of cells over time, since the data we have refer to different aliquots and enriching processes. We can nevertheless normalise the numbers to the HSCs and pool the mice over time among the mice, also because the ratio is the observable considered in the model through this work anyway: Our aim is to use the hypothesis of steady state for the total number of cells in a population, thus such number should be constant, or, similarly, the slope of the linear regression should be compatible with zero. We report the 95% confidence bound in the square brackets below each plot in the above figure. We see that the steady state assumption is solid for all compartments but DP and the B cells in the bone marrow, for which, however, the maximum increase is less than 2 fold. We thus consider them constant too.

The test on the ratios does not guarantee that all sizes are constant over time. We checked for a few time points whether the number of HSC in a comparable amount of bone marrow are changing over time. We see that the number of HSC starts increasing when a mouse is older than 250 days, so after the duration of the experiment. We nevertheless tested the compatibility with our model to one in which HSC increases (and thus, since the ratios are about constant, so do the other compartments). We found no contradiction in the estimated kinetic parameters, although the confidence bound in this case are larger due to the larger errors on the pooled compartment ratios (Figure 8.2).

2.3.1

It has been suggested that HSCs are heterogeneous and biased to independently differentiate into their progeny according to different routes. One
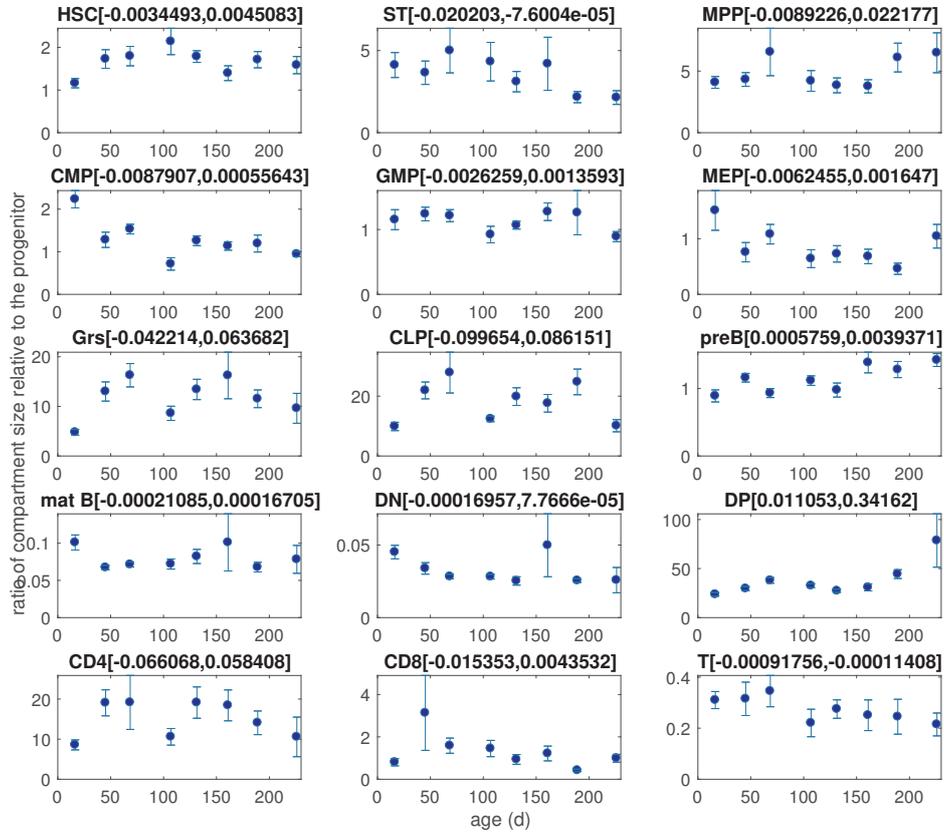
Figure 8.1: Time course of the pooled ratio of all populations with respect to HSC. Square brackets represent 95% confidence interval on the best fit value for the slope of the linear regression.
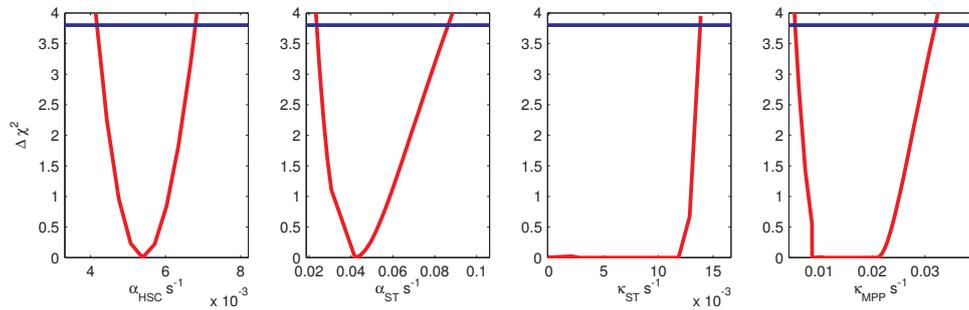
Figure 8.2: Parameters' profiles in case the number of HSCs increases over time.
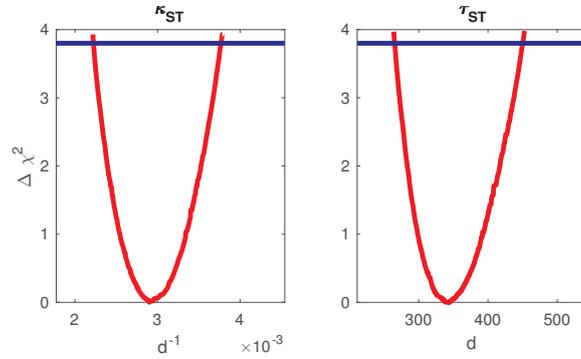
Figure 8.3: Profile likelihood of net efflux and residence time in ST-HSC population.
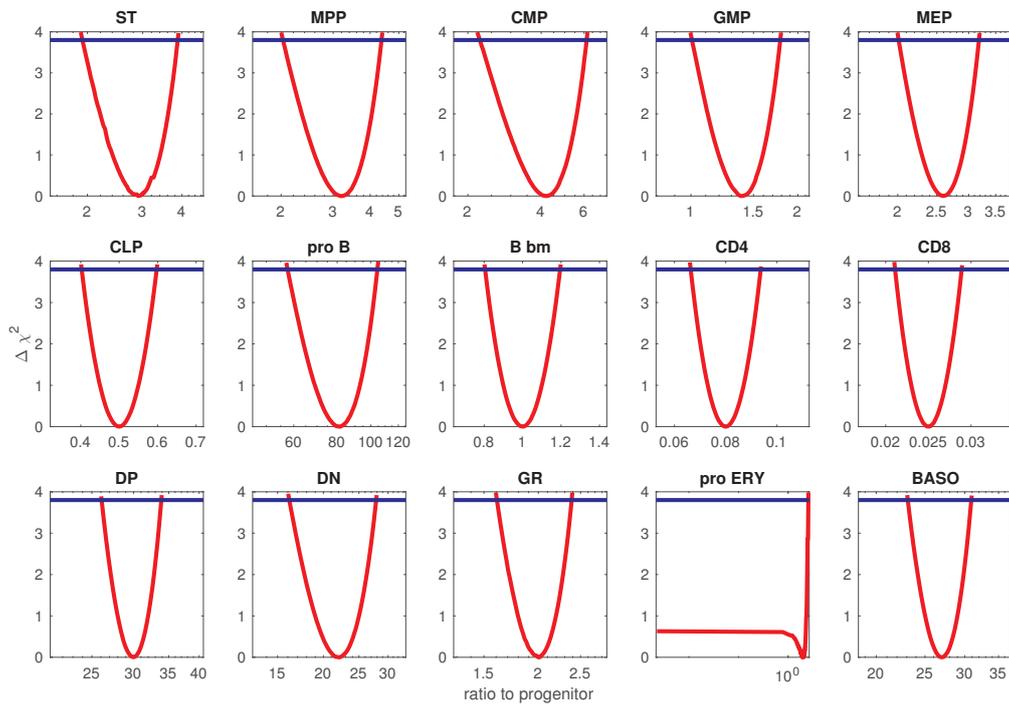
Figure 8.4: Profile likelihood of the ratio of compartment size with respect to their progenitor.

possible scheme to describe this possibility would be in this case: Now, if
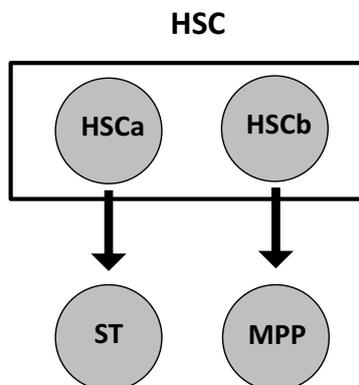
Figure 8.5: Possible scheme to account for different differentiation pathways among the HSC population.

the initial labelling frequency is different for the subpopulations $HSC_{a,b}$, over time we would see the labelling frequency in ST-HSC and MPP equilibrating to $HSC_{a,b}$ respectively. Since the observed labelling frequency in HSC is lower than one among $HSC_{a,b}$, the labelling frequency in one progeny will overshoot the one in HSC at some point in time.

2.2

	κ (d^{-1})	τ (d)	α (d^{-1})	ratio to HSC
HSC	0 (assumed)	inf (assumed)	0.009 [0.005 0.012]	1
ST	0.003 [0.002 0.004]	340 [260 450]	0.07 [0.03 0.3]	3 [2 4]
MPP	0.02 [0.01 0.08]	50 [10 80]	4 [0.7 4] in CMP 0.008 [0.005 0.017] in CLP	9 [5 15]

Table 8.1: Estimated parameters for the stem cells

3.2

Let us assume that the fraction of active cells in a certain stem population is less than one. In this case we would have quiescent cells, meaning cells that do nothing in steady state at all time points. The fraction of BrdU positive cells would evolve over time according to:

$$\dot{B}^+(t) = (2\,\sigma_a + \gamma_a)\,(F_a - B^+(t)) = \nu_a\,(F_a - B^+(t)) \tag{8.1}$$

	$\kappa\ (d^{-1})$	$\tau\ (d)$	$\alpha\ (d^{-1})$	ratio to HSC
CMP	1 [0.6 2]	1 [0.5 6]	3 [0.2 4] in GMP 1.6 [0.1 1.6] in MEP	4 [2 7] ·10
GMP	2 [0.14 3.5]	0.6 [0.3 7]	4 [0.3 4]	5 [2 10] ·10
MEP	0.4 [0.04 2]	2 [0.6 22]	1 [∼0 4]	10 [6 19] ·10
GR s	2 [0.14 2.5]	0.5 [0.4 7]	-	10 [4 20] ·10
pro Ery	0.12 [0.03 4]	8 [0.3 40]	1[0.5 4]	3 [0 5] ·10^3
baso	0.04 [0.02 4]	30 [6 60]	-	1 [0 40] ·10^4

Table 8.2: Estimated parameters for the myeloid cellls

	$\kappa\ (d^{-1})$	$\tau\ (d)$	$\alpha\ (d^{-1})$	ratio to HSC
CLP	0.017 [0.009 0.035]	6 [3 10] *10	1.5 [0.6 3] in pro B 1 [0.33 3] in DN	5 [2 8]
pro B	0.05 [0.007 0.078]	2 [6 12]	0.06 [0.03 0.07]	4 [2 7] ·10^2
B	0.051 [0.012 4]	20 [0.3 80]	-	4 [2 7] ·10^2
DN	0.02 [0.014 4]	20 [6 60]	4 [0.7 4]	1 [5 2] ·10^2
DP	0.13 [0.02 4]	70 [6 50]	0.01 [0.001 0.4] in CD4 0.002 [0.0003 0.1] in CD8	3 [1 5] ·10^3
CD4	1 [0.01 4]	2 [0.3 80]	-	2 [1 4] ·10^2
CD8	4 [0.01 4]	13 [0.3 80]	-	8 [4 14] ·10

Table 8.3: Estimated parameters for the lymphoid cells

where "a" stands for active, σ_a and γ_a are the rates at which active cells symmetrically or asymmetrically divide respectively, $\nu_a = 2\,\sigma_a + \gamma_a$ is the rate of self-renewal of active cells and F_a is the fraction of active cells in the population.

On the other hand, the equation we presented in section 3.2 reads:

$$\dot{B}^+(t) = (2\,\sigma + \gamma)\,(1 - B^+(t)) = \nu\,(1 - B^+(t)) \qquad (8.2)$$

where σ and γ are the rates at which cells symmetrically or asymmetrically divide respectively assuming that all cells are active and $\nu = 2\,\sigma + \gamma$. Let us choose ν_a in order to properly describe the data for HSC-1. We plot the average self-renewal rate $\nu_r = F_a\,\nu_a$ against ν as inferred upon fitting Equation 8.2 to Equation 8.1 for all possible values of the fraction of active cells: We see that already for an active fraction of around 30% the estimated self-
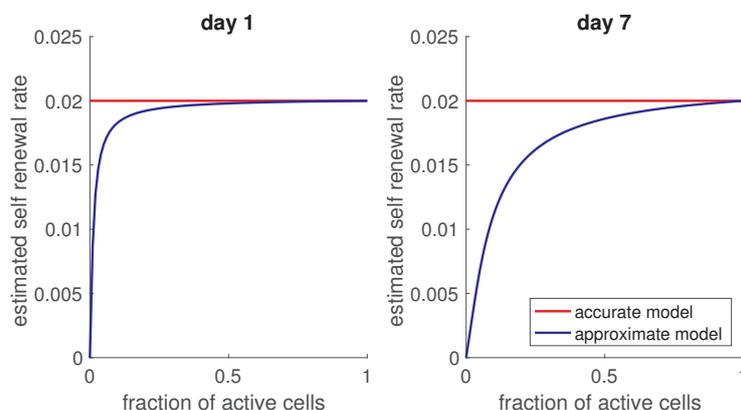
Figure 8.6: Comparing a model that allows for some cells not being active (accurate model) to model that assumes they all are (approximate model). The estimated average self renewing rates of the stem population are comparable if the fraction of active cells is at least 30%.

renewal rate are similar, comparably to our result for the limited dilution analysis 2.1.5.

The possibility of having exclusively active cells is consistent with [Cheshier et al., 1999], who observed that after 6 months of BRDU administration all HCSs are BrdU$^+$, but not with [Bernitz et al., 2016], according to whose experiment a small fraction of HSCs never divide in a life span of a mouse.

We now compare the black box model to the phase-resolved model:
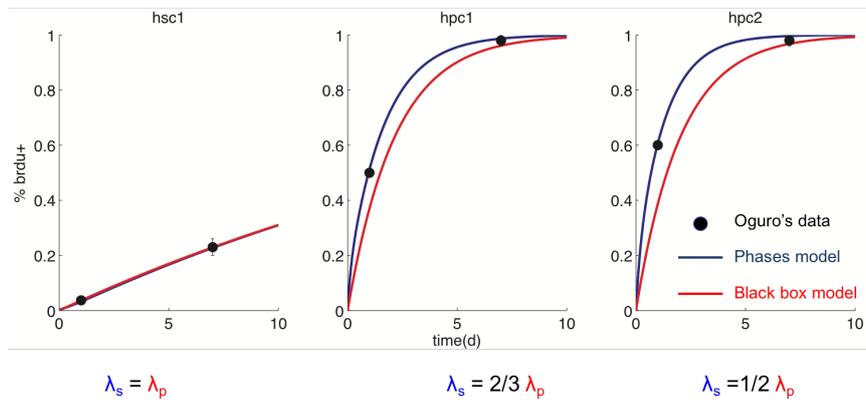
Figure 8.7: For the more mature populations (HPC-1,2), estimates of the proliferation rate with the black box model is less accurate than taking the phases resolution into account.

Abbreviations

HSCs	haematopoietic stem cells
ST-HSCs	short term haematopoietic stem cells
MPPs	multipotent progenitors
CMPs	common myeloid progenitors
GMPs	granulocyte-monocyte progenitors
MEPs	megakaryocyte-erythroid progenitors
GRs	granulocytes
CLPs	common lymphoid progenitors
proBs	B cells progenitors
DNs	double negative cells
DPs	double positive cells
CD4s	CD4 positive T cells
CD8s	CD8 positive T cells
proERYs	progenitors of erythroid cells
baso	basophils
MCM	improved Cre fused with two modified oestrogen receptors
YFP	yellow fluorescent protein
bm	bone marrow

Bibliography

[Abkowitz et al., 2000] Abkowitz, J. L., Golinelli, D., Harrison, D. E., and Guttorp, P. (2000). In vivo kinetics of murine hemopoietic stem cells. *Blood*, 96(10):3399–3405.

[Akashi et al., 2000] Akashi, K., Traver, D., Miyamoto, T., and Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774):193–197.

[B and (1993)., ages] B, E. and (1993)., T. R. (New York, 436 pages.). An introduction to the bootstrap.

[Benz et al., 2012] Benz, C., Copley, M. R., Kent, D. G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., Treloar, D., Brinkman, R. R., and Eaves, C. J. (2012). Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell*, 10(3):273–283.

[Bernitz et al., 2016] Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H., and Moore, K. (2016). Hematopoietic stem cells count and remember self-renewal divisions. *Cell*, 167(5):1296–1309.e10.

[Boggs, 1984] Boggs, D. R. (1984). The total marrow mass of the mouse: A simplified method of measurement. *American Journal of Hematology*, 16(3):277–286.

[Boyer et al., 2010] Boyer, S. W., Schroeder, A. V., Smith-Berdan, S., and Forsberg, E. C. (2010). All hematopoietic cells develop from hematopoietic stem cells through flk2/flt3-positive progenitor cells. *Cell Stem Cell*, 9(1):64–73.

[Bradford, 1997] Bradford, Williams, R. B. (1997). Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. *Exp Hematol.*

[Buchholz et al., 2013] Buchholz, V. R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Gräf, P., Verschoor, A., Schiemann, M.,

Höfer, T., and Busch, D. H. (2013). Disparate individual fates compose robust cd8+ t cell immunity. *Science*, 340(6132):630–635.

[Burnham and Anderson., 2002] Burnham, K. P. and Anderson., D. R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. *Springer, 2 edition.*

[Busch et al., 2015] Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., and Rodewald, H.-R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 518:542 EP –.

[Chen et al., 2016] Chen, J. Y., Miyanishi, M., Wang, S. K., Yamazaki, S., Sinha, R., Kao, K. S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I. L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. 530:223 EP –.

[Cheshier et al., 1999] Cheshier, S. H., Morrison, S. J., Liao, X., and Weissman, I. L. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proceedings of the National Academy of Sciences*, 96(6):3120–3125.

[Dick et al., 1985] Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A., and Bernstein, A. (1985). Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of w/wv mice. *Cell*, 42(1):71 – 79.

[Dingli et al., 2007a] Dingli, D., Traulsen, A., and Michor, F. (2007a). (a)symmetric stem cell replication and cancer. *PLOS Computational Biology*.

[Dingli et al., 2007b] Dingli, D., Traulsen, A., and Pacheco, J. M. (2007b). Compartmental architecture and dynamics of hematopoiesis. *PLOS ONE*, 2(4):1–4.

[Ema et al., 2005] Ema, H., Sudo, K., Seita, J., Matsubara, A., Morita, Y., Osawa, M., Takatsu, K., Takaki, S., and Nakauchi, H. (2005). Quantification of self-renewal capacity in single hematopoietic stem cells from normal and lnk-deficient mice. *Developmental Cell*, 8(6):907 – 914.

[Flossdorf, 2013] Flossdorf, M. (2013). Stochastic t cell fate decisions. *PhD thesis*.

[Ford et al., 1956] Ford, C. E., Hamerton, J. L., Barnes, D. W. H., and Loutit, J. F. (1956). Cytological identification of radiation-chimæras. *Nature*, 177:452 EP –.

[Foudi et al., 2008] Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J. W., Jaenisch, R., Carey, V., and Hock, H. (2008). Analysis of histone 2b-gfp retention reveals slowly cycling hematopoietic stem cells. *Nature Biotechnology*, 27:84 EP –.

[Ganusov and Boer, 2012] Ganusov, V. V. and Boer, R. J. D. (2012). A mechanistic model for bromodeoxyuridine dilution naturally explains labelling data of self-renewing t cell populations. *J R Soc Interface*.

[Gerrits et al., 2010] Gerrits, A., Dykstra, B., Kalmykowa, O. J., Klauke, K., Verovskaya, E., Broekhuis, M. J. C., de Haan, G., and Bystrykh, L. V. (2010). Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood*, 115(13):2610–2618.

[Gillespie, 1976] Gillespie, D. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *JOURNAL OF COMPUTATIONAL PHYSICS*.

[Golub and Cumano, 2013] Golub, R. and Cumano, A. (2013). Embryonic hematopoiesis. *Blood Cells, Molecules, and Diseases*, 51(4):226 – 231. Embryonic and Fetal Hematopoiesis.

[Guo et al., 2013] Guo, G., Luc, S., Marco, E., Lin, T.-W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G.-C., and Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, 13(4):492–505.

[Hoppe et al., 2016] Hoppe, P. S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K. D., Hilsenbeck, O., Moritz, N., Endele, M., Filipczyk, A., Gambardella, A., Ahmed, N., Etzrodt, M., Coutu, D. L., Rieger, M. A., Marr, C., Strasser, M. K., Schauberger, B., Burtscher, I., Ermakova, O., Bürger, A., Lickert, H., Nerlov, C., Theis, F. J., and Schroeder, T. (2016). Early myeloid lineage choice is not initiated by random pu.1 to gata1 protein ratios. *Nature*, 535(7611):299–302.

[Ito and Ito, 2016] Ito, K. and Ito, K. (2016). Self-renewal of a purified tie2+. *Science*.

[Jordan and Lemischka, 1990] Jordan and Lemischka (1990). Clonal and systemic analysis of long.term hematopoiesis in the mouse. *Genes and Development*.

[Karamitros et al., 2018] Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., Usukhbayar, B., Calvo, J., Taylor, S., Goardon, N., Six, E., Pflumio, F., Porcher, C., Majeti, R., Göttgens, B., and Vyas, P. (2018).

Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nature Immunology*, 19(1):85–97.

[Kay, 1965] Kay, H. (1965). How many cell-generations ? *The Lancet*, 286(7409):418 – 419. Originally published as Volume 2, Issue 7409.

[Keller et al., 1985] Keller, G., Paige, C., Gilboa, E., and Wagner, E. F. (1985). Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. *Nature*, 318:149 EP –.

[Keller and Snodgrass, 1990] Keller, G. and Snodgrass, R. (1990). Life span of multipotential hematopoietic stem cells in vivo. *Journal of Experimental Medicine*, 171(5):1407–1418.

[Kiel et al., 2007] Kiel, M. J., He, S., Ashkenazi, R., Gentry, S. N., Teta, M., Kushner, J. A., Jackson, T. L., and Morrison, S. J. (2007). Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain brdu. *Nature*, 449:238 EP –.

[Kiel et al., 1995] Kiel, M. J., Yilmaz, Ö. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., and Morrison, S. J. (1995). Slam family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*, 121(7):1109–1121.

[Kondo et al., 1997] Kondo, M., Weissman, I. L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, 91(5):661–672.

[Lai and Kondo, 2006] Lai, A. Y. and Kondo, M. (2006). Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Journal of Experimental Medicine*, 203(8):1867–1873.

[Laurenti and Dick, 2012] Laurenti, E. and Dick, J. E. (2012). The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature Immunology*.

[Lemischka et al., 1986] Lemischka, I. R., Raulet, D. H., and Mulligan, R. C. (1986). Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell*, 45(6):917 – 927.

[Li and Slayton, 2013] Li, X. and Slayton, W. B. (2013). Molecular mechanisms of platelet and stem cell rebound after 5-fluorouracil treatment. *Experimental Hematology*, 41(7):635 – 645.e3.

[Lu et al., 2011] Lu, R., Neff, N. F., Quake, S. R., and Weissman, I. L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nature Biotechnology*, 29:928 EP –.

[Luchsinger et al., 2016] Luchsinger, L. L., de Almeida, M. J., Corrigan, D. J., Mumau, M., and Snoeck, H.-W. (2016). Mitofusin 2 maintains haematopoietic stem cells with extensive lymphoid potential. *Nature*, 529:528 EP –.

[MacArthur and Lemischka, 2013] MacArthur, B. D. and Lemischka, I. R. (2013). Statistical mechanics of pluripotency. *Cell*, 154(3):484–489.

[Mackey, 2001] Mackey, M. C. (2001). Cell kinetic status of haematopoietic stem cells. *Cell Proliferation*, 34(2):71–83.

[Manesso et al., 2013] Manesso, E., Teles, J., Bryder, D., and Peterson, C. (2013). Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation. *Journal of The Royal Society Interface*, 10(80).

[Marciniak-Czochra et al., 2008] Marciniak-Czochra, A., Stiehl, T., Ho, A. D., Jäger, W., and Wagner, W. (2008). Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation. *Stem Cells and Development*, 18(3):377–386.

[Martins et al., 2012] Martins, V. C., Ruggiero, E., Schlenner, S. M., Madan, V., Schmidt, M., Fink, P. J., von Kalle, C., and Rodewald, H.-R. (2012). Thymus-autonomous t cell development in the absence of progenitor import. *J Exp Med*.

[Mohri et al., 1998] Mohri, H., Bonhoeffer, S., Monard, S., Perelson, A. S., and Ho, D. D. (1998). Rapid turnover of t lymphocytes in siv-infected rhesus macaques. *Science*, 279(5354):1223–1227.

[Morrison and Kimble, 2006] Morrison, S. J. and Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441:1068 EP –.

[Morrison and Weissman, 1994] Morrison, S. J. and Weissman, I. L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*, 1(8):661–673.

[Naik et al., 2013] Naik, S. H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R. J., and Schumacher, T. N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*, 496:229 EP –.

[Notta et al., 2016] Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y., and Dick, J. E. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, 351(6269).

[Nowell et al., 1956] Nowell, P. C., Cole, L. J., Habermeyer, J. G., and Roan, P. L. (1956). Growth and continued function of rat marrow cells in x-radiated mice. *Cancer Research*, 16(3):258–261.

[Nutt et al., 2005] Nutt, S. L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of pu.1 expression in multipotent hematopoietic progenitors. *Journal of Experimental Medicine*, 201(2):221–231.

[Oguro et al., 2013] Oguro, H., Ding, L., and Morrison, S. J. (2013). Slam family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cells*.

[Pei et al., 2017] 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., and Rodewald, H.-R. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, 548:456 EP –.

[Perié et al., 2015] Perié, L., Duffy, K. R., Kok, L., de Boer, R. J., and Schumacher, T. N. (2015). The branching point in erythro-myeloid differentiation. *Cell*, 163(7):1655–1662.

[Perié et al., 2014] Perié, L., Hodgkin, P. D., Naik, S. H., Schumacher, T. N., de Boer, R. J., and Duffy, K. R. (2014). Determining lineage pathways from cellular barcoding experiments. *Cell Reports*, 6(4):617 – 624.

[Pietras et al., 2015] Pietras, E. M., Reynaud, D., Kang, Y.-A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. M., Göttgens, B., and Passegué, E. (2015). Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell*, 17(1):35–46.

[Pohjanpalo, 1978] Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 41(1):21 – 33.

[Rauch et al., 2009] Rauch, M., Tussiwand, R., Bosco, N., and Rolink, A. G. (2009). Crucial role for baff-baff-r signaling in the survival and maintenance of mature b cells. *PLoS ONE*, 4(5):e5456.

[Raue et al., 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

[Rodriguez-Fraticelli et al., 2018] Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., and Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, pages EP –.

[Sawai et al., 2016] Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D. J. H. F., Lavin, Y., Lau, C. M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., Mirny, L. A., Merad, M., Eaves, C. J., and Reizis, B. (2016). Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. *Immunity*, 45(3):597–609.

[Schittler et al., 2013] Schittler, D., Allgöwer, F., and De Boer, R. J. (2013). A new model to simulate and analyze proliferating cell populations in brdu labeling experiments. *BMC Systems Biology*, 7(1):S4.

[Schoedel et al., 2016] Schoedel, K. B., Morcos, M. N. F., Zerjatke, T., Roeder, I., Grinenko, T., Voehringer, D., Göthert, J. R., Waskow, C., Roers, A., and Gerbaulet, A. (2016). The bulk of the hematopoietic stem cell population is dispensable for murine steady-state and stress hematopoiesis. *Blood*, 128(19):2285–2296.

[Shin et al., 2014] Shin, J.-W., Buxboim, A., Spinler, K. R., Swift, J., Christian, D. A., Hunter, C. A., Léon, C., Gachet, C., Dingal, P. C. D. P., Ivanovska, I. L., Rehfeldt, F., Chasis, J. A., and Discher, D. E. (2014). Contractile forces sustain and polarize hematopoiesis from stem and progenitor cells. *Cell Stem Cell*, 14(1):81–93.

[Spalding et al., 2005] Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H., and Frisén, J. (2005). Retrospective birth dating of cells in humans. *Cell*, 122(1):133 – 143.

[Stumpf et al., 2017] Stumpf, P. S., Smith, R. C. G., Lenz, M., Schuppert, A., Müller, F.-J., Babtie, A., Chan, T. E., Stumpf, M. P. H., Please, C. P., Howison, S. D., Arai, F., and MacArthur, B. D. (2017). Stem cell differentiation as a non-markov stochastic process. *Cell Systems*, 5(3):268–282.e7.

[Sun et al., 2014] Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature*, 514:322 EP –.

[Takizawa et al., 2011] Takizawa, H., Regoes, R. R., Boddupalli, C. S., Bonhoeffer, S., and Manz, M. G. (2011). Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. *Journal of Experimental Medicine*, 208(2):273–284.

[Tough and Sprent, 1994] Tough, D. F. and Sprent, J. (1994). Turnover of naive- and memory-phenotype t cells. *Journal of Experimental Medicine*, 179(4):1127–1135.

[Vácha et al., 1980] Vácha, J., Holá, J., Dungel, J., and Znojil, V. (1980). The volume of plasma and erythrocytes in individual bones and in the spleen of mice under physiological conditions and with acute radiation-induced atrophy of the haemopoietic tissue. *Acta Haematologica*, 64(3):165–171.

[van der Wath RC, 2009] van der Wath RC, Wilson A, L. E. T. A. L. P. (2009). Estimating dormant and active hematopoietic stem cell kinetics through extensive modeling of bromodeoxyuridine label-retaining cell dynamics. *PLoS ONE*.

[van Furth R, 1989] van Furth R (1989). Origin and turnover of monocytes and macrophages. *Curr Top Pathol*.

[Wilson et al., 2008] 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., and Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, 135(6):1118–1129.

[Wu et al., 2007] Wu, M., Kwon, H. Y., Rattis, F., Blum, J., Zhao, C., Ashkenazi, R., Jackson, T. L., Gaiano, N., Oliver, T., and Reya, T. (2007). Imaging hematopoietic precursor division in real time. *Cell Stem Cell*, 1(5):541–554.

[Yamamoto et al., 2013] Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*, 154(5):1112–1126.

[Yang et al., 2005] Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Månsson, R., Sigvardsson, M., and Jacobsen, S. E. W. (2005). Identification of lin–sca1+kit+cd34+flt3– short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood*, 105(7):2717–2723.