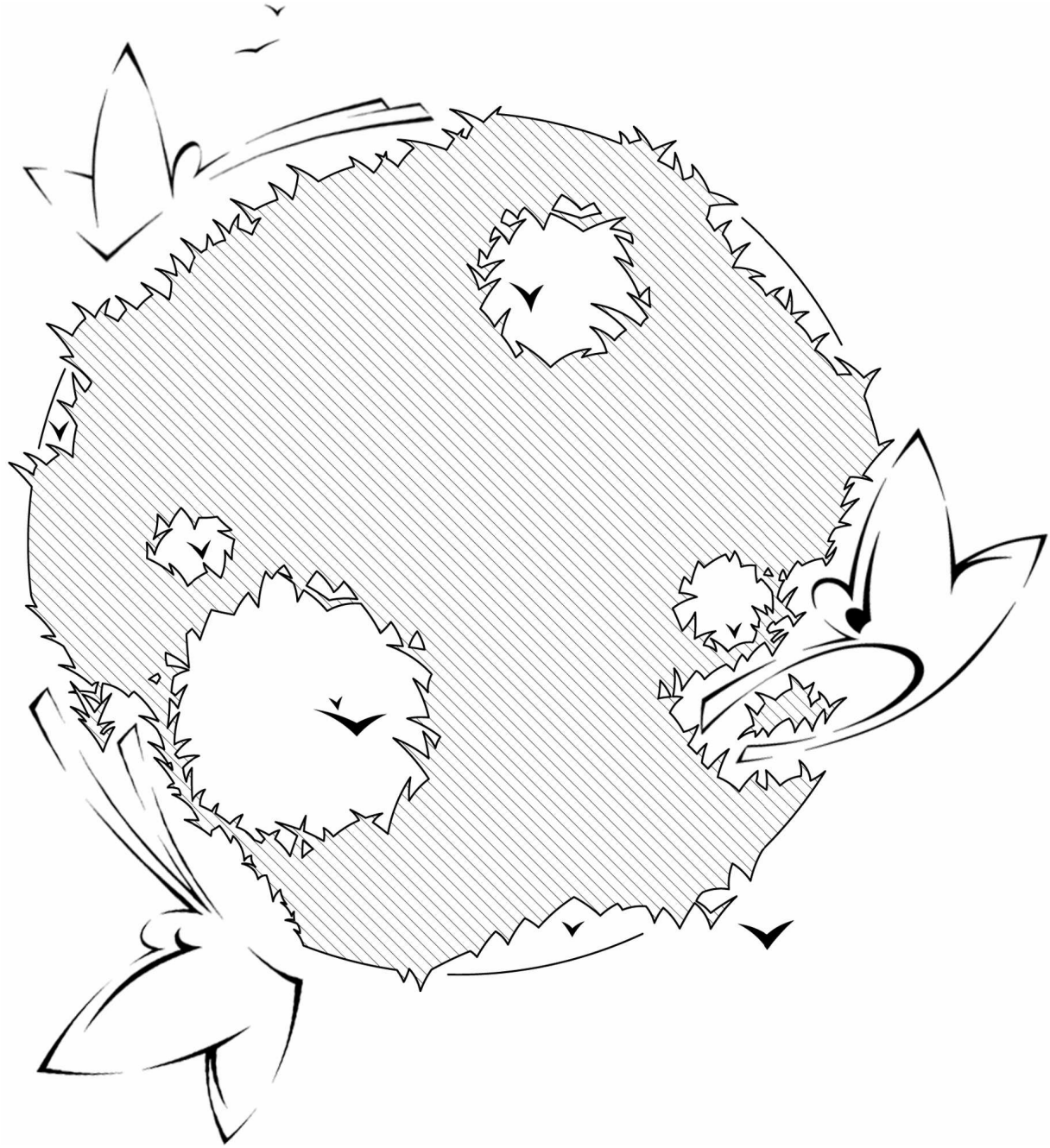

Theoretical studies on the lineage specification of hematopoietic stem cells

Ingmar Glauche

Ph.D. thesis

University of Leipzig, Germany, 2010

To my daughters

Elisabeth Katharina and Henriette Marie

Contents

Denken und Wissen sollten immer gleichen Schritt halten.
Das Wissen bleibt sonst tot und unfruchtbar.

Wilhelm von Humboldt
(1767 - 1835)

1. Introduction and objective

1.1. State of research

Stem cells and their therapeutic applications. Tissue stem cells are found in all multi-cellular organisms that retain the ability to renew themselves through mitotic cell division. In order to maintain or reconstitute a particular tissue these cells can specialize into a wide range of mature cell types through a process of differentiation and lineage specification. Given the developmental potential of stem cells, it is not surprising that the interest in their therapeutical use is continuously growing.

A distinction between tissue stem cells and embryonic stem cells is appropriate in the first place. Tissue stem cells, which are also called adult or somatic stem cells, are typically found within a particular tissue or organ. There, the tissue stem cells guarantee the sustained function of the tissue by the maintenance of their own population and by the generation of the specialized cell types of the tissue which are needed for replacement and repair [1, 2]. Generally, their differentiation potential is narrowed to the cell types of the particular tissue. In contrast, embryonic stem cells are derived from the cells of the inner cell mass of a blastocyst and are maintained in *in vitro* cultures [3, 4, 5]. These cells are defined by their ability to differentiate into all cell types of the organism. However, they represent a fixation of a transient cell stage (corresponding to the cells of the inner cell mass of the blastocyst) and therefore do not have a stable *in vivo* counterpart. Although embryonic stem cells have a superior differentiation potential compared to tissue stem cells, their application for medical therapy imposes a number of technical problems and ethical questions which are more extensively discussed in Section 2.6 of this thesis.

A complete new perspective in experimental stem cell research was recently opened by the observation that somatic cells can be reprogrammed to a state closely similar to embryonic stem cells [6, 7, 8, 9] thus circumventing nuclear transfer techniques. Although the approaches are based on the targeted overexpression of a set of relevant genes using gene vectors, the results gained great attention. In the very first place, these results showed that a reprogramming of a complete somatic cell is in principle possible. Additionally, these techniques do not require the utilization of human egg cells and the generated cells are of the same genotype as the donor ones. These advantages, to circumvent the ethical discussion using embryonic stem cells or human eggs as well as the identity with the genotype of the donor, put again highest expectations on stem cell research. Potential ap-

plications in the clinical setting depend very much on the understanding of the underlying process to successfully tune the reprogramming and the subsequent differentiation.

So far, only tissue stem cells are successfully used for medical treatment or are at least studied in the clinical situation. The most prominent example is the use of hematopoietic stem cells (HSCs) for the treatment of severe malignancies of the blood generating system. As in most leukemias normal hematopoiesis is displaced by a rapidly expanding clone of malignant cells with impaired function. Current transplantation treatments aim on an eradication of the complete hematopoietic system of the effected patient and the subsequent replacement of the HSC source provided from a healthy donor. Given the successful engraftment of the donor stem cells in the bone marrow of the host they can completely reconstitute the hematopoietic system and provide all types blood cells as there are lymphocytes, erythrocytes and granulocytes. However, there is still a number of uncertainties with stem cell transplantation, e.g. concerning the chances of donor engraftment and the risk of immunoreactions between host and donor tissue.

Further feasible applications of stem cells include the use of epidermal stem cells and the resulting keratinocytes in a wide spectrum of different techniques for the tissue-engineering of artificial skin (see [10] and references therein). Such cultured skin substitutes are regularly used for burn-injured patients and for the treatment of chronical wounds. By reducing the demand for autologous skin transplants the cultured skin substitutes have significantly enhanced the survival rates for major burns and reduced the efforts for skin biopsies in patients with chronic wounds.

Novel therapeutic approaches have identified tissue stem cells as possible targets for the insertion of genes which are corrupted or deficient in patients suffering from severe genetic, often inheritable malignancies. This approach, commonly called *gene therapy*, was successfully applied for the treatment of Severe Combined Immunodeficiency Disorder (SCID) [11], a further subform called X-SCID [12] and for the treatment of Chronic Granulomatous Disease (CGD) [13]. These studies showed that the integration of the missing or corrupted gene in HSCs using viral vectors can contribute to the corrected tissue function over extended time periods. However, since the integration site of the corrected gene is random, the transfected cell sample might contain cells in which other functions are corrupted and potentially induce tumors. Such effects have to account for a number of secondary hematopoietic malignancies with lethal consequences [14] and are the greatest limitation to the present application of gene therapy.

Current limitations and novel concepts. The limited number of available medical applications is a clear indicator that many questions about general principles of stem cell organization, culturing and differentiation are not fully answered. Even from the perspective of basic research many details about the intrinsic reg-

ulation of stem cells, their interactions with the local environment, their differentiation decisions and even their identification are still unknown.

The limitations in the understanding of basic principles became even more obvious when the classical view on tissue stem cells was challenged by experimental observations demonstrating the ability of such cells to contribute to functional cell types of completely different tissues. It has for example been shown that HSCs have the general ability to differentiate into heart tissue ([15], see also the critical discussion in [16]), liver [17], brain [18, 19] and muscle [20, 21]. This phenomenon, generally referred to as *plasticity*, has again increased the interest in tissue stem cells since it offers an alternative approach to circumvent the ethically questionable use of embryonic stem cells. Even more important, these findings challenge to whole concept of perceiving tissue stem cells as a rather fixed and homogeneous population of cells with well defined characteristics and predetermined differentiation potential. It appears that these stem cells are much more flexible than previously anticipated and that they are generally able to respond to changing environmental conditions.

This is not only true for the unexpected plasticity of these cells to switch between tissue, but also for their population-inherent heterogeneity. It could be shown that stem cell populations are rather heterogeneous with respect to the cycling status or with respect to the expression of surface markers [22, 23]. It is therefore not surprising, that a prospective identification of a HSC could not be accomplished so far. Even the most rigorous isolation protocols based on the expression of characteristic surface markers result in heterogeneous populations that are enriched for HSCs but in which a significant number of cells fail to demonstrate long-term reconstitution of a depleted animal [24]. For the prospective identification of stem cells this isolation strategy might not be the ultimate tool since the flexible and reversible expression of the surface markers as well as their limited functional correlation to the stem cell activity are principle objections. But also on the underlying genetic level it is not yet clear whether there exist a unique stem cell signature or not [25]. Even more, it might be just the defining characteristic of stem cells to *not* express a strict pattern uniquely encoding pluripotency and the potential for sustained self-renewal. It seems much more plausible that the stem cell state is characterized by different facets of flexibility, reversibility and robustness, and that this state is highly dependent on the environmental conditions, defined by the tissue, the signaling context and the metabolic options. In this sense "stemness" is not a specific cellular property but rather an inherent function of a particular heterogeneous cell population.

In the same line of argument it is questionable whether the process of lineage specification, which describes the transition of originally undifferentiated (stem) cells into functional mature cell types, is a predetermined sequence of decision steps or whether lineage specification needs to be understood as a flexible and adaptive process that is only predictable on the population level. As early as

in the 1980s, Ogawa and coworkers [26, 27, 28] studied the lineage specification potential of hematopoietic progenitor cells and concluded that committed cells are derived from multipotent progenitors through a progressive restriction of lineage potential, in which the restriction in type and number of lineages occurs in a stochastic fashion. Although many details on the molecular mechanisms of lineage specification have been elucidated since then it remains unclear how the specific gene expression dynamics are generated and how they are controlled by cell-cell and cell-environment interactions. Even on a more general level it is not well understood how processes like adhesion, migration and apoptosis influence the cell fate development. However, there is an increasing number of reports that agree upon a general phenomenology of the lineage specification process: whereas the undifferentiated state is characterized by a low level coexpression of many lineage specific and potentially antagonistic genes in early progenitor cells, this coexpression, commonly referred to as *priming*, disappears in the course of differentiation, when certain lineage restricted genes are up-regulated while others are down-regulated [29, 30, 31, 32].

Theoretical approaches in stem cell biology. The above criticism is not in place to discredit more than 40 years of successful research on stem cells. It is the motivation to study stem cell organization in a comprehensive, conceptual fashion in which different and diverse phenomena can be uniquely represented and linked to each other. It is evident, that every experimental design requires at least a conceptual perception of the underlying biological processes in order to either embed the results in the existing framework or to modify or even falsify the existing model. Prominent examples of such conceptual models are the hierarchy of hematopoietic cell differentiation or the interaction maps that outline the interplay of genes and transcription factors. However, beyond the level of conceptual models is the level of mathematical validation which represents a more rigorous description of the conceptual framework in terms of quantitatively accessible measures and predictions.

There is little argument that stem cell organization is a complex regulation process emerging from the collective interaction of a multitude of components. However, complexity is not just a "game of many players" that need to be identified. One could even argue that the identification of the players has nothing to do with complexity at all but it is the interactions, i.e. "the rules of the game", that make up complexity. In this case, it is not sufficient to analyze rather isolated components or pathways since the organizational principles only appear in the context of many interaction partners. This perception leads to an understanding in which determinism is replaced by the notion of dynamically stabilized systems. It is here that the plasticity and reversibility of cellular features that are controlled by interactions with the environment guarantee for a robust and adaptive

organization. Such systems are generally referred to as *self-organizing*.

The small number of quantitative modeling approaches in the field of stem cell biology does not compare to the importance of this research field. Besides the obstacles in identification and measurement of this rare cell class, many models are rather descriptive in nature and stick to the traditional categories of stem cell organization as a deterministic process. However, there is also few quantitative models that apply the idea of complex interacting systems to the stem cell field and that contributed to the development of novel conceptual ideas about stem cell organization [1, 33, 34]. Within such models it is possible to consistently describe a wide range of experimental findings and to couple them to a perspective which perceives stem cell organization as an adaptive and flexible process.

New challenges in stem cell research. Within the last years the systems biological advances made the field of stem cell biology more attractive for novel theoretical approaches. Especially the availability of vast amounts of molecular data such as mRNA and protein screens as well as the rapid increase of computational resources fostered the development of many new analytical methods for data mining and representation. Driven by the experimental research many such approaches aim at the identification of molecular key players that influence stem cell development and organization. Translating these findings from the molecular level to the phenotypic level is still a major challenge.

Apart from the molecular data, the availability of new monitoring techniques such as high-resolution time lapse video microscopy facilitates the observation of cell cultures over time. Analyzing the development of individual stem cells and their progeny within such cultures yields information about the divisional history, the timing of differentiation events like asymmetric cell fates and the role of cellular interactions. Regarding each cellular trajectory as a possible realization of the developmental potential of the original (stem) cell, a multitude of such trajectories should provide an impression about the diversity of possible developments. Such approaches are suited to couple observed phenomena on the population level to the relevant processes on the cellular level. However, the interpretation and generalization of such cellular trajectories requires a rigorous analytical foundation, which is in many parts still lacking.

Model approaches in stem cell biology are important tools to connect cell intrinsic regulations like signaling pathways and gene expression patterns on one side with the population behavior on the other side. The scope of such models aims at the identification of general principles that define the appearance of the tissue, e.g the regulatory principles leading to self-maintenance of the stem cell population, the system inherent heterogeneity or the mechanisms that govern the differentiation process into different types of functional cells. Given the necessary simplifications from the rather complex biological system, these models are in-

adequate for the understanding of molecular details. However, such abstractions are often key to understanding the fundamental properties of such complex systems and allow to put different experimentally observed phenomena in a common conceptual and quantitatively accessible context. This helps to identify common regulatory principles, and therefore in the design of future experimental strategies.

1.2. Motivation

Phenomenology. Undifferentiated stem cells are defined by their ability to differentiate into distinct somatic cell types. For the hematopoietic system, HSCs can generally contribute to red and different types of white blood cells. This differentiation process, generally involving a sequence of decision steps, is referred to as *lineage specification* and is characterized by a number of biologically relevant features:

- **Loss of lineage potential.** In a seminal series of experiments in the 1980s, Ogawa and coworkers [26, 27, 28, 35] studied the lineage specification potential of hematopoietic stem and progenitor cells. It could be demonstrated that the potential for the contribution to multiple lineages (multipotency) is progressively lost while the cells undergo differentiation. This lead to the conclusion that lineage specification is characterized as a progressive restriction of lineage potential until a distinct mature cell fate is finally realized.

- **Generation of diversity.** Analyzing single cell-derived colonies in a number of closely similar studies [26, 27, 36] it could be furthermore shown that originally undifferentiated cells can give rise to heterogeneous colonies containing multiple different cell types. Together with the above mentioned loss of multi-lineage potential this generation of diversity is commonly associated with a "hierarchy of stem cell differentiation". In this conceptual picture the stem cell at the origin gives rise to all lineage restricted progenitors and mature cells in a hierarchic sequence of decision steps (compare Figure 2.2).

- **Regulation of lineage specification.** Lineage specification is not a static process but can be regulated both *in vivo* and *in vitro*. Common examples include the adaptation to higher demands of erythrocytes in high-altitude conditions [37] or the targeted differentiation of primary cells and cell lines *in vitro* (see e.g. [38, 39]).

- **Temporal extension.** The process of lineage specification has a temporal extension. As illustrated e.g. in [39, 40] the proportion of undifferentiated cells decreases continuously as a function of time while the proportions of functional cells increases over the same period.

- **Reversibility.** Generally it appears that reversibility plays a minor role under homeostatic conditions but might be essential in cases of injury and repair. Experimental evidence suggests that lineage specification is at least a partially reversible process [22, 41, 42].

- **Priming.** On the molecular level, the undifferentiated state is characterized by a coexpression of many, potentially antagonistic genes and transcription factors [29, 31, 43]. This low-level coexpression, commonly referred to as *priming*, disappears in the course of differentiation when certain lineage-restricted genes are up-regulated while others are down-regulated.

- **Lineage bias.** Experimental evidence suggests that HSCs can maintain an inheritable lineage bias in the sense that they preferentially contribute to one cell type as compared to another [44, 45]. The mechanisms introducing this additional level of heterogeneity among multipotent stem and progenitor cells are still speculative.

For further details of the discussed phenomena the reader is referred to Chapter 2 of this work.

Conceptual approach. Besides their ability to provide a multitude of distinct functional cell types for replacement and repair, stem cells are furthermore characterized by their ability to preserve the regenerative potential of a tissue by sustained maintenance of their stem cell population. To guarantee both these functions certain mechanisms are required that on one hand regulate the self-maintenance of the stem cell population, even under changing demands, and on the other hand generate a stable, although tunable, composition of different functionally differentiated cells.

Adopting a critical and unprejudiced view on stem cell organization it seems obvious to ask whether such a system can be envisioned in the context of self-organization and reactive adaptiveness, in which stemness is not a cellular property of single cells but a systemic feature. Approaching this subject from the perspective of a theoretical biologist, these questions have to be addressed in a well defined and mathematically founded fashion. On this level it needs to be asked which general types of rules are required to establish a self-organizing system satisfying the experimentally imposed criteria. In turn, these general principles have implications for the interpretation of the system behavior and even more direct impact on measurable experimental outcomes.

For the mathematical description of the sustained maintenance of the stem cell population, an elegant approach has been advocated by Ingo Roeder and Markus Loeffler [33, 46, 47]. This approach, which is established for the well studied hematopoietic system, is build on the idea that each cells is subject to different

environmental signals which in turn induce different directions for the developments of these cells. For a simplified system with just two opposing signaling contexts, one inducing differentiation and loss of stem cell function, the other inducing self-maintenance and regeneration, it could be shown that the major experimental results can be reproduced as there are the sustained maintenance of the HSC population, the dynamic response to system depletion, the heterogeneity of the stem cell population and the establishment of cellular chimerisms. A detailed account of the model is given in Section 3.4. However, the model by Roeder and Loeffler does not account for the second functionality of tissue stem cells: the sustained, although adaptive supply of different types of mature cells for maintenance and repair of the functional tissue.

1.3. Objective

It is the objective of this work to establish a theoretical model which accurately represents the phenomenology of stem cell differentiation and lineage specification and to study its implications on the interpretation of cellular behavior both on the single cell as well as on the population level. In particular, the following issues are addressed in this thesis:

- A generalized analytical framework has to be developed to understand lineage specification as an intracellular, temporally extended process based on the interaction of competitive "differentiation programs". This concept needs to be discussed on the basis of a generalized understanding of the molecular process involved in lineage specification of HSCs.

- The proposed intracellular lineage specification dynamics have to be embedded in the model for HSC organization proposed by Roeder and Loeffler [33]. To do so, the interpretation of an instructive microenvironment has to be extended to the situation of lineage specification, thus inducing a correlation between the regulation of self-renewal and lineage specification.

- The intracellular model of lineage specification needs to represent different possible mechanisms for the regulation of lineage contribution (i.e. *instructive* and *selective* lineage specification).

- To verify the proposed theoretical model, the simulation results need to be compared to several sets of experimental data covering different aspects of the lineage specification process.

In the second part of the thesis the extended model approach is applied to the tracing of individual cell fates including the cell's complete progeny. Therefore, the following aspects have to be covered:

- Establishment of a topological characterization for the tree-like structures resulting from the single cell-based approach. This needs to be supplemented by the proposition of suitable measures for their analysis and comparison.
- The application of these measures and their robustness needs to be demonstrated under different (artificial) conditions. In particular, it has to be addressed how different modes of lineage specification can be inferred from the single cell tracking data.

The thesis is structured as follows. Chapter 2 provides a broad overview of the experimental notion of stem cells and illustrates in more detail the phenomenology of lineage specification. Chapter 3 continues with an introduction to the relevant conceptual and mathematical models. Based on a catalogue of criteria a theoretical concept of lineage specification is proposed in Chapter 4 and complemented by a sensitivity analysis and a comparison with experimental data in Chapter 5. Extending the focus towards the analysis of single cell fates Chapter 6 introduces the relevant nomenclature and proposes appropriate measures. Chapter 7 analyses the suitability of these measures for different *in silico* situations. The thesis is concluded with a comprehensive summary and outlook in Chapter 8.

2. Biological background: Stem cells and lineage specification

2.1. Stem cells: history and definitions

It was as early as in 1868 that the biologist Ernst Haeckel used the term "Stammzelle" (german for *stem cell*) to refer to the unicellular ancestor from which multicellular organisms developed during evolution. By mapping such an evolutionary "Stammbaum" (german for *pedigree* or *family tree*) onto an embryological tree he advocated the view that the founding fertilized egg can also be called "stem cell". It was around the turn of that century that the term "stem cell" was also used to identify a common progenitor cell that gives rise to different cell types of the peripheral blood. However, the notion of one common type of stem cell that regenerates the complete hematopoietic system remained controversial. It was not until the early 1960s that Canadian scientists around Ernest McCulloch and James Till could prove the existence of a common undifferentiated hematopoietic (stem) cell that is capable of self-renewal and differentiation. In a series of experiments it could be shown that single, undifferentiated bone marrow cells yield the potential to developed into spleen colonies of irradiated mice [48, 49, 50]. An excellent review about the historical usage of the term "stem cell" was published recently by Ramalho-Santos and Willenbring [51].

After the landmark experiments by Till and McCulloch modern stem cell research took up speed. Using transplantations experiments a whole body of knowledge about the repopulation potential of certain bone marrow derived cells, their differentiation hierarchy as well as their contribution to multiple hematopoietic lineages has been established. Beyond the hematopoietic system, somatic stem cells could also be identified in other tissues, e.g. in the skin [52], gut [53], liver [54], neural tissue [55]. The apparent proximity between the processes of somatic stem cell differentiation and differentiation during embryonic development established the research on early embryonic cells as an additional branch of stem cell research. After the first stable embryonic stem cell line has be derived from the inner cell mass of a mouse embryo in 1981 [3, 4], this research quickly expanded, resulting in the first human embryonic stem cell line derived in 1997 [56]. In contrast to tissue stem cells, embryonic stem cells are defined *in vitro* and are characterized by their ability to self-renew in culture and to contribute to all tissues of the three primary germ layers, ectoderm, endoderm and mesoderm. In 2006, Shinya Yamanaka and colleagues showed the possibility to reprogram tissue

cells into an embryonic stem cell-like state by targeted overexpression of four specific transcription factors Sox-2, Oct-4, c-Myc, and Klf4 [6, 8]. The availability of these so called induced pluripotent stem (iPS) cells opened a completely new perspective for stem cell research and its applications in the 21nd century.

These recent advances are also based on the rapid biotechnical developments in the last decades that opened the window for the understanding of molecular processes in stem and progenitor cells. Using these techniques it was possible to identify many important genes that are required for stem cell maintenance and differentiation as well as their regulation by transcription factors and signaling pathways. Especially the availability of microarrays, that could identify the whole range of mRNAs transcribed within a certain cell population, raised the hope for a unique "transcriptional signature" of stem cells [57, 58]. However, such a unique signature could not be obtained so far [25], and it is questionable whether such an stable signature exists or if the stem cell state is dynamically defined [59]. Similarly, much effort has been made to prospectively isolate stem cells based on the expression of certain surface antigens. But also these techniques only enrich for stem cells and it is not yet possible to prospectively decide whether a particular cell will act as a stem cell in a functional assay or not. Therefore, the "gold standard" for defining a stem cell is a functional characterization evaluating the cell's ability for long-term tissue regeneration *in vivo*.

There is an abundance of definitions that commonly characterize stem cells by their ability to maintain their own population (self-maintenance) as well as their ability to contribute to different types of functional cells (multipotency). However, such working definitions do mostly not account for the full spectrum of phenomena associated with stem cells nor do they clearly outline whether the assigned features apply to isolated cells or to a cell population. Potten and Loeffler [1], as well as other scientists [60, 61] illustrated that stem cell function needs to be understood as a concerted action of many different cells including stem cells and their local environment. This in turn implies that a functional definition of stem cell is not a characterization of a single cell but of a heterogeneous population. Aiming on a comprehensive representation of the full functional spectrum of tissue stem cells and based on a previous work [1], Loeffler and Roeder proposed a more sophisticated definition of stem cells [46]. They define tissue stem cells as a potentially heterogeneous population of functionally undifferentiated cells, capable of:

- homing to an appropriate growth environment
- proliferation
- production of a large number of differentiated progeny

- self-renewing or self-maintaining their own population
- regeneration of the functional tissue after injury with
- flexibility and reversibility in the use of these options.

A detailed account of the characteristic features is provided in the original publication [46], however some aspects are of particular relevance for the focus of this thesis and are briefly discussed below. First, the definition is based on "capabilities" which is a linguistic synonymous of "potential": certain functional tasks can be accomplished *in principle*. Whether the potential is actually used, is highly dependent on the particular situation (e.g. homeostasis, injury). Here, and second, another aspect comes into play, which is the role of the particular environment. It could be shown for different tissues that stem cell regulation is highly dependent on the supportive function of the local environment which is commonly referred to as "niche". Within these niches stem cells are subject to different environmental signals, which are generally supportive in nature and allow them to adaptively react on changing demands. The fact that it has not been possible to expand HSCs *in vitro* is most likely due to the incomplete understanding of the complex signaling that occurs in the *in vivo* niches.

Third, the production of a large number of differentiated progeny includes the option for unipotent stem cells that only give rise to one mature cell type, like certain stem cells of the epidermis. Forth, the terms self-renewal and self-maintenance are often used interchangeable, although they might refer to slightly different phenomena [62]. Self-maintenance generally refers to the property of a stem cell population to preserve their functionality to act as stem cells, which is the general case in a homeostatic situation. In contrast, self-renewal does also include the repopulation of a tissue and the reestablishment of the corresponding stem cell population after serious stress. However, it is more philosophical questions, whether these are really distinct processes and if just one of both is sufficient to fulfill the definition criteria. Fifth, there is a strong emphasize on the flexibility in the use of the assigned capabilities. Since the list of criteria postulates a number of different potentials, a particular cell does not necessarily have to make use of *all* of them. It might well be the case that under certain environmental conditions just one of the capabilities is used (e.g. the differentiation into a large number of functional cells) whereas other capabilities (e.g. the self-maintenance) remain unused. It might even be the case that the use of the potentials changes with time including reversible developments. At this point it is most evident that any functional definition of stem cells generally refers to a population of such cells. Since the criteria are functional characteristics, a single stem cell cannot fulfill them all at the same time. In this respect it is appropriate to abandon the view of stem cell entities and to refer to the *stemness* of cell populations.

2.2. Hematopoietic stem cells and hematopoiesis

Hematopoietic stem cells are responsible for the production of blood cells throughout the life of an organism. The origin of these cells during embryonic development is not completely elucidated. However, it could be demonstrated that the formation of hematopoietic stem cells occurs in a complex developmental process that involves several anatomical sites including the yolk sac, the placenta and the fetal liver [63]. Finally, at birth, a pool of HSCs is established in the bone marrow [64], in mice especially in the femur, in humans also in the pelvis, sternum and vertebra. The bone marrow contains a particular micro-environment which on one hand supports the maintenance of the stem cell population and on the other hand provides a surplus of cells for differentiation and the production of mature blood cells. In the post-natal, homeostatic situation, HSCs are generally in a quiescent state and are only rarely activated into cell cycle [65, 66, 67].

Hematopoietic stem cells are a rare subpopulation among the cells in the bone marrow. Estimates in mice revealed that between one to eight HSCs can be found within 10^5 nucleated bone marrow cells [68, 69]. However, HSCs are morphologically undistinguishable from other cells in the bone marrow which do not have the stem cell functionality. Although, the expression of certain surface markers has successfully been used to enrich for HSCs up to a purity of one stem cell in two selected cells it is still not possible to prospectively identify a stem cell with certainty [70].

Unlike other stem cell systems with a well-structured spatial arrangement (e.g. the small intestine or the hair follicle), the identification of HSCs based on spatial information is not possible. It is the semi-liquid composition of the bone marrow that makes it very challenging to study HSCs and their interactions with the local micro-environment in the *in vivo* situation [71]. This is also the reason why a precise identification of the hematopoietic niche as a complex of local and systemic growth factors, stroma cells, and extra-cellular matrix components, is such a difficult task. However, it is a well established fact that the interaction of stem cells with is local environment are indispensable for the functionality. Originally, the perception of an inductive microenvironment was introduced in early 1970s by John Trentin [72, 73], however it is generally Raymond Schofield [74] who is credited with the concept of a stem cell niche. In this interpretation a niche corresponds to a defined anatomical site that regulates stem cell function by means of secreted and cell surface molecules, mechanical signals, spatial arrangements and particular metabolic conditions [75, 76]. Although precise knowledge about the spatial structures is missing, a schematic visualization of the concept is provided in Figure 2.1.

Meticulous work revealed that osteoblast cells play a crucial role in the construction of the hematopoietic niches. Osteoblasts potentially mediate the adhesion of HSCs to the particular niches by the formation of N-cadherin/β-catenin adherens

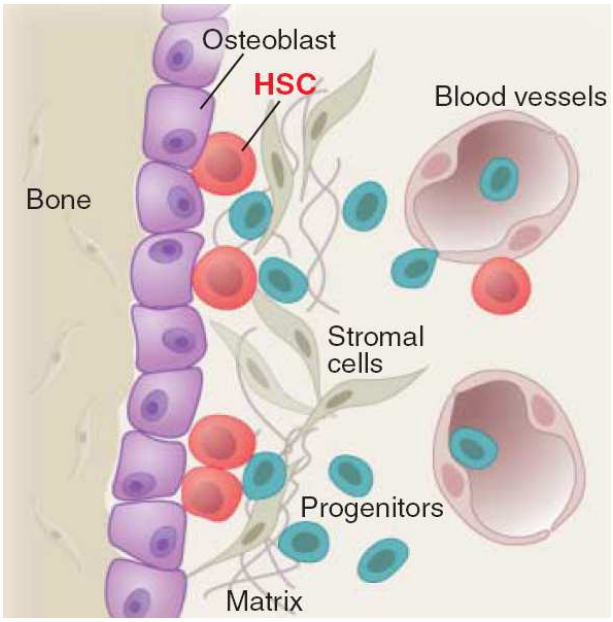

Figure 2.1.: The hematopoietic niche.
In the picture, stem cells are shown in close proximity to the osteoblast and sheltered by stromal cells and extracellular matrix components. Picture reproduced from [77] with kind permission from Kateri Moore.

complexes. N-cadherin function is decreased in the course of differentiation which in turn might promote the displacement from osteoblastic niche. A major role to facilitate a connection between osteoblasts and HSCs has been demonstrated for the Ang-1/Tie2 signaling, which is important for sustained quiescence and long-term repopulation ability [78]. Similar regulating functions have been assigned to the Notch and Wnt pathways, however the exact regulation needs further investigation. An excellent overview about the state of research has been published in a review article by Moore and Lemischka [77].

The availability of osteoblast niches seems to be a major regulator of the absolute stem cell number [79]. However, besides the dominance of the osteoblast niches in the bone marrow also other tissues, like spleen and liver, can provide similar functionality [80, 81] and might replace the bone marrow niches in situations of stress [82]. In this context, it has recently been shown that also sinusoidal endothelial cells in bone marrow establish an alternative, so called "vascular" niche for HSCs [83]. Whereas in the osteoblast niches stem cells are primarily maintained in the quiescent state, vascular niches seem to be involved in the regulation of stem cell proliferation, differentiation and mobilization. However, the mutual regulation between these different supportive environments is only insufficiently understood.

A surplus of cells that is not necessary for the maintenance of the stem cell pool undergoes differentiation. These cells progressively lose their potential to act as stem cells, undergo massive expansion and develop into mature cell types that are finally released into the peripheral blood. Specifically, three different types (lineages) of blood cells are all derived from a common class of HSCs, namely erythrocytes, leukocytes and platelets. *Erythrocytes* are responsible for the oxygen transport from the lungs to the different organs and tissues. This transport is facilitated by the protein hemoglobin, which is uniquely present in erythrocytes and the corresponding precursor cell stages. Furthermore, erythrocytes are characterized by the loss of their nucleus which has been discharged during the late normoblast stage. In humans, erythrocytes circulate in the peripheral blood for about 120 days. *Leukocytes* are composed of a mixture of different cell types, most of which are primarily involved in the immunoreactions of the organism. The mutual interaction of these cells is important for an optimal immuno-defense, although the different leukocytes differ significantly in their appearance and function. It is generally distinguished between granulocytes (neutrophils, eosinophils, and basophils), lymphocytes and monocytes / macrophages. Leukocytes persist in the peripheral blood between a few hours (neutrophils) up to months and years (lymphocytes). *Thrombocytes*, also known as platelets, are responsible for blood coagulation and wound closure. The average life span for thrombocytes is about 10 days.

The intermediate developmental stages leading to the production of the mature blood cells are often comprised in a hierarchical tree structure. Besides the final, mature cell types many precursor stages can be identified based on morphology or the expression of characteristic surface markers. The particular processes that regulate the development of the distinct cell types are generally referred to as *lineage specification*. Since it is the focus of the thesis to provide a principle understanding of lineage specification, this subject is more extensively discussed in the Section 2.3.

In the light of the definition of stem cells provided in the previous section, HSCs are a primary example to fit the stated criteria. For this system, transplantation experiments have been established as the most rigorous proof of stem cell function since all criteria need to be fulfilled in order to guarantee long-term reconstitution. In such experiments, donor-derived HSCs are transplanted in lethally irradiated animals that do not have functional hematopoiesis anymore. It has been shown that already one HSC is sufficient to reestablish complete hematopoiesis to normal levels and to rescue the animal [84]. However, the application of repopulation experiments also has serious limitations. E.g. the process of irradiation causes a number of side effects that are hard to quantify. Furthermore, the reestablishment of hematopoiesis requires the support of a functioning micro-environment but also the contribution of other blood cells to guarantee the animal's survival briefly after transplantation.

2.3. Lineage specification of hematopoietic stem cells

Besides the developmental options of self-renewal, differentiation, migration and apoptosis, stem cells and their progenitors undergo a decision process to select the type of functional cells to which they contribute. The question, how stem cells specify this cell type, is long-standing and will be elucidated from a conceptual perspective within this thesis.

Throughout the thesis the term *lineage specification* is used to characterize the sequence of decision steps that finally leads to lineage commitment, manifested by the acquisition of particular lineage specific features. The term *differentiation* refers to the actual process of acquisition of these lineage specific features and does also include the loss of self-renewal ability. *Lineage potential* describes the general ability of a stem or progenitor cell to produce a certain number of different types of mature functional cells. In contrast, the *lineage contribution* of a particular cell describes the cell types actually produced in a particular differentiation process, and may therefore comprise only a part of the cell's lineage potential.

2.3.1. Phenomenology

The hematopoietic system is commonly perceived as a hierarchy with the stem cell at the origin giving rise to lineage restricted progenitors and finally to terminally differentiated end cells [85, 86, 87, 88], see also Figure 2.2. Different developmental stages within in this tree-like differentiation sequence have been characterized as follows: The pool of multipotent HSCs consists of long-term reconstituting hematopoietic stem cells (LT-HSCs) that can repopulate an lethally irradiated animal and establish life-long supply of all hematopoietic lineages. Furthermore, the multipotent HSCs also include the so called short-term repopulating hematopoietic stem cells (ST-HSCs) that also give rise to all hematopoietic lineages, but do so for only 8-10 weeks. It is commonly assumed that ST-HSCs are an intermediate cell stage between LT-HSCs and the more lineage restricted progenitors as they are the common myeloid progenitor (CMP) and the common lymphoid progenitor (CLP). In a further set of decisions CLPs give rise to the different cells of the lymphatic lineages: B- and T-lymphocytes and natural killer cells. In contrast, CMP normally differentiate to cells of the myeloid lineages, among them granulocytes, macrophages, erythrocytes and thrombocytes.

The hematopoietic system is able to propagate changes in the demand of peripheral blood cells using feedback mechanisms that act on earlier cell stages. Such regulations are facilitated e.g. by the release of cytokines and chemokines [89, 90]. By adjustments of early lineage decisions and increased expansion the system is able to adaptively respond to such changing needs which are introduced e.g. by injuries, infections or high altitudes.

In a series of experiments in the 1980s, Ogawa and coworkers [27, 28, 35] studied

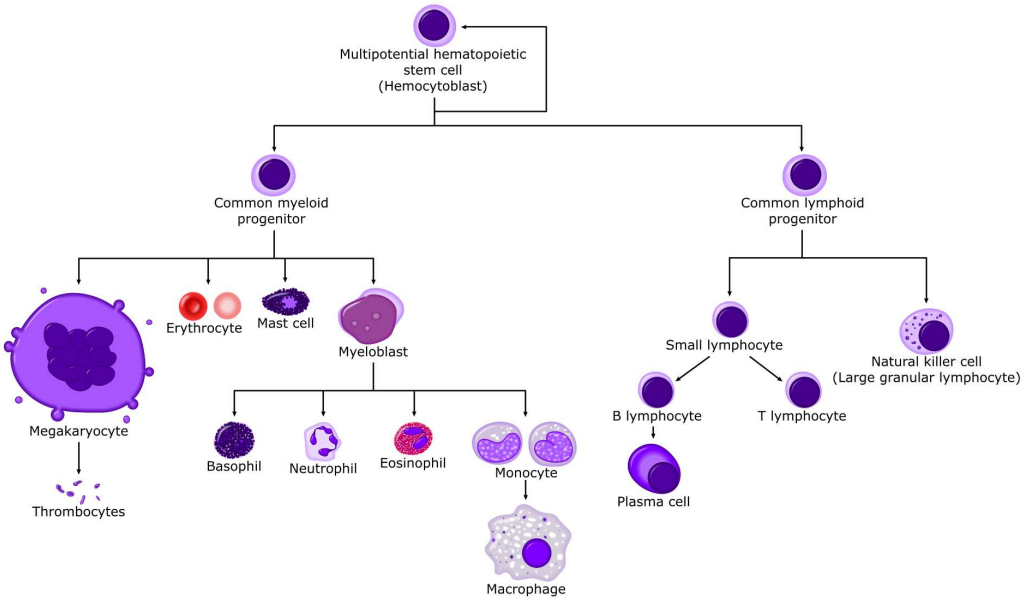

Figure 2.2.: The hematopoietic system.
Different, prominent cell types of the hematopoietic system are arranged in a "classical" hierarchical ordering that represents the major lines of development. Picture taken from [91].

the lineage specification potential of hematopoietic progenitor cells. In particular they used hematopoietic, spleen-derived mouse cells [27] and human umbilical cord blood cells [28] to examine and compare the developmental fate of two daughter cells derived from one parent cell. From their results the authors concluded that committed cells are derived from multipotent progenitors through a progressive restriction of lineage potential, in which the restriction in type and number of lineages occurs in a stochastic fashion. In their interpretation, which still holds today, the contribution of lineages which is observed in the progeny of single differentiating cell can only be predicted in a statistical sense. The particular experiments will serve as a primary reference to verify the proposed modeling approach.

The question remains, how the phenomenological view on lineage specification can be mapped on an appropriate molecular basis in which lineage specification is interpreted as a shift in gene expression patterns and correlated to epigenetic modifications.

2.3.2. Molecular aspects of lineage specification

Although many details on the molecular mechanisms of lineage specification are still unknown, it appears that certain molecular processes facilitate lineage specification of multipotent progenitor cells to select and to commit to one out of a finite number of predefined lineages. Since lineage committed, mature cells are characterized by discrete and robust "genetic programs" supporting the functional and morphological requirements, lineage specification is the process of establishing the particular cell type-specific gene expression pattern [92]. The processes involved do not necessarily act on the transcriptional level alone but potentially include all kinds of molecular regulations, e.g. post-transcriptional and translational modifications, phosphorylations, signaling pathways and epigenetic alterations. Because of this complexity of possible interactions a comprehensive, conceptual understanding of the nature of multipotency and the dynamic processes that initiate and regulate lineage specification is still incomplete.

Priming. Although great effort has been made to obtain characteristic transcriptional profiles of various stem cells [57, 58, 93]), no consensus of a unique, stable stem cell state could be reached. It becomes increasingly evident that multipotency, as one of the defining characteristics of stem cells, is a dynamically stabilized feature associated with the simultaneous coexpression of lineage specific, potentially antagonistic genes [29, 31, 93, 43, 59]. Such a "promiscuous" coexpression of genes specific for alternative lineages fates, commonly referred to as *priming*, has been reported for multipotent cell types of the hematopoietic system [30, 32]. It can be envisioned as the molecular representation of the potential for different developmental options which disappears in the course of differentiation when certain lineage restricted genes are up-regulated while others are down-regulated. A sketch is provided in Figure 2.3.

Regulation of lineage specification. The regulation of specification is ultimately required to respond to changing demands of mature blood cells. Such control mechanisms are commonly facilitated by the action of regulatory molecules, such as cytokines, chemokines and growth factors like erythropoietin or G-CSF [89, 90] that mediate changing demands and skew the lineage specification process. For the action of such external signals two general modes of regulation are discussed [94, 95, 87, 96]: In the *instructive* mode the external signals do directly impose their regulation on the transcriptional level by promoting or repressing certain gene activities via signal transduction cascades. This means that the cell-intrinsic lineage decision is skewed. In contrast, in the *selective* mode, the external signals lack lineage-determining capacity and the cells randomly decide between preexisting developmental fates. This way, the cell intrinsic decision is not skewed and the external signals exert their action by promoting the survival of already

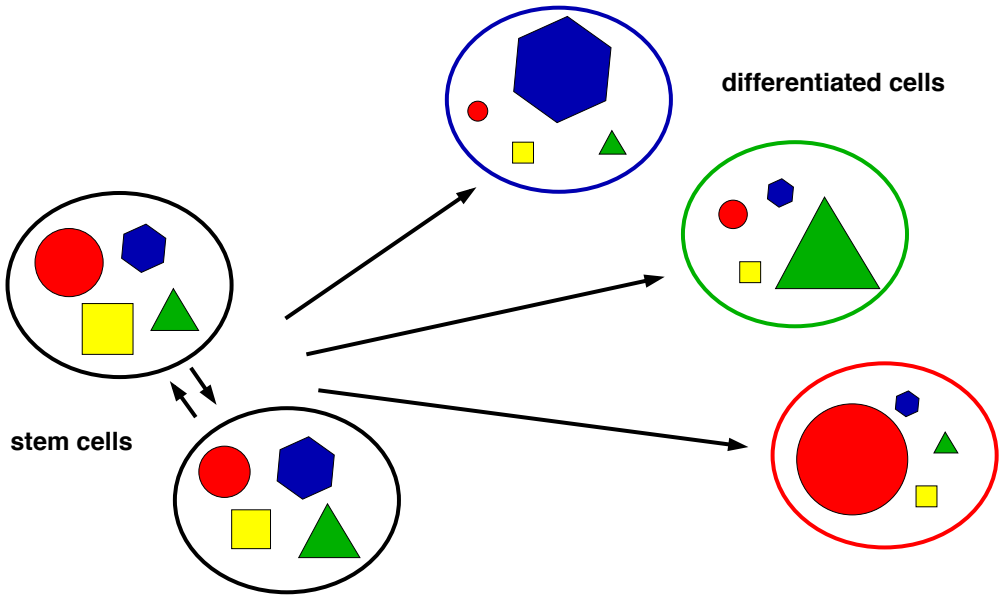

Figure 2.3.: Priming.
The *priming* behavior of the HSCs is depicted by the equal sized symbols indicating lineage specific gene expression patterns. In the course of differentiation one lineage specific program is finally up-regulated while others are down-regulated. This change in the expression patterns is encoded in the symbol sizes and the corresponding "color-coding" of the cell membrane.

committed cells. In particular it appears that certain cytokines support the survival of cells expressing a subset of lineage specific receptors while in the absence of the cytokines the cells undergo cell death. Using such a regulation it is possible to skew an initially balanced cell population to preferentially contribute to some but not all available lineages. These modes of lineage specification are briefly outlined in Figure 2.4 in which the lineage decision process is separated from the survival under appropriate conditions.

The latter, *selective* mode of lineage specification is often referred to as stochastic (or intrinsic) as compared to deterministic (or extrinsic) which is also used as an alternative description for the *instructive* mode. However, this terminology is confusing and not suited to distinguish properly between these modes of lineage specification. Perceiving the cell-intrinsic regulation in the *instructive* mode as a shift of probabilities (rather than the deterministic on-off mechanism) for the different available cell fates, the actual cell fate decisions can still be stochastic. Therefore, the more discriminative terms *instructive* (regulated cell-intrinsic decision) and *selective* (unregulated cell-intrinsic decision and selective survival) will be used throughout this thesis.

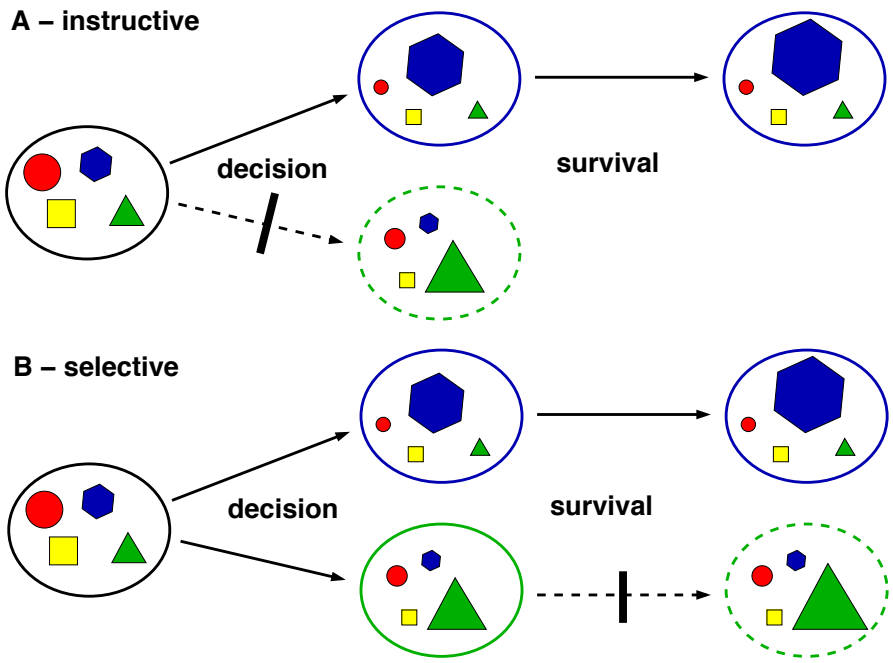

Figure 2.4.: Modes of lineage specification.
(A). In the *instructive* mode of lineage specification the lineage decision is intrinsically skewed to commit to one cell type instead of another (indicated by the block). Selective survival plays a minor role. **(B).** In the *selective* mode the intrinsic lineage decision is rather balanced giving rise to cells of different lineages. However, external signals selectively promote the survival of some lineages whereas others are discarded (indicated by the block).

Temporal extension. The process of lineage specification is not realized "all-of-a sudden" but involves a temporal extension in which the cells pass through several intermediate stages (see Figure 2.5). On the molecular level, this refers to the time which is necessary to remodel the genetic expression pattern including the corresponding changes on the epigenetic level. Furthermore, the expression pattern needs to be stabilized by the activation of down-stream target genes and the subsequent changes in cellular morphology.

In this concept, in which lineage specification is understood as a temporally extended sequence of molecular alterations, it is immediately evident that an understanding of the underlying process requires the measurement of time courses of the development. Such time structured experiments that analyze the phenotypical changes of cell cultures are rather well established [39, 97], although homogeneous cell populations are still difficult to achieve. The availability of high-throughput methods allows to extend these approaches towards the molecular level [40, 98]. Nevertheless, these experiments suffer from a structural insufficiency of stem cell

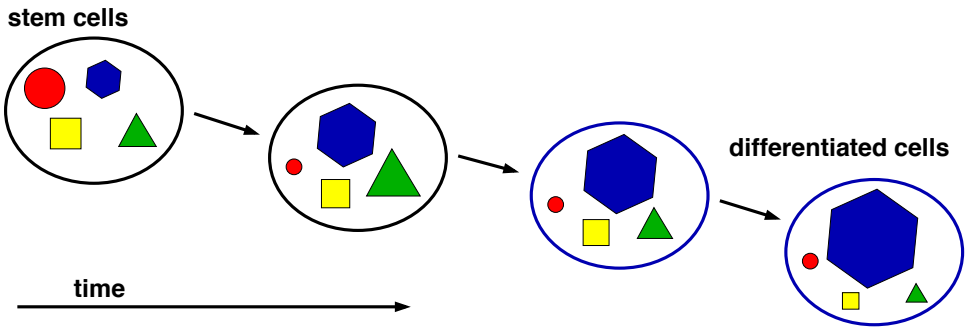

Figure 2.5.: Temporal extension.
Lineage specification is described as temporally extended process in which a cell transits through different intermediate steps. Starting from *primed* stem cells the up-regulation of a lineage specific gene expression patterns involves different stages of remodeling, indicated by the changing sizes of the colored symbols.

cultures since they are based on cell populations rather then on individual cells. Given the inherent heterogeneity of such cultures the experimental findings are always averages of distinct cells at potentially different developmental stages and can hardly be attributed to a particular cell. However, for the example of the PU.1-GATA-1 interaction on the level of common myeloid progenitors (CMP), the high-throughput data has been used to analyze the dynamics of the molecular switch in a time dependent manner [92].

Key regulators and switches. It could be shown successfully that a number of transcription factors are actively involved in the process of lineage specification [43, 99, 100, 101]. Although a comprehensive understanding of their interactions is still missing, it has been proposed that such key regulators act on particular "switch points" of the developmental sequence by activation of a certain lineage specific program at the cost of an alternative program [64]. Figure 2.6 provides a sketch illustrating the typical cross-antagonism between two generic transcription factors.

As a prominent example, it has been shown on the level of common myeloid progenitors (CMP) that the overexpression of transcription factor GATA-1 can induce differentiation towards erythroid/megakaryocyte development as opposed to myeloid lineages [102]. A similar mechanism is know for transcription factor PU.1 although in the opposite direction [103]. Such a fate switch between two alternative lineage-specific "programs" can be achieved if the key regulator for one lineage program has a direct negative feedback on the competing lineage programs. In the case of transcription factors PU.1 and GATA-1 it could be shown that besides promoting their "own" program (autoregulation) the transcription factors

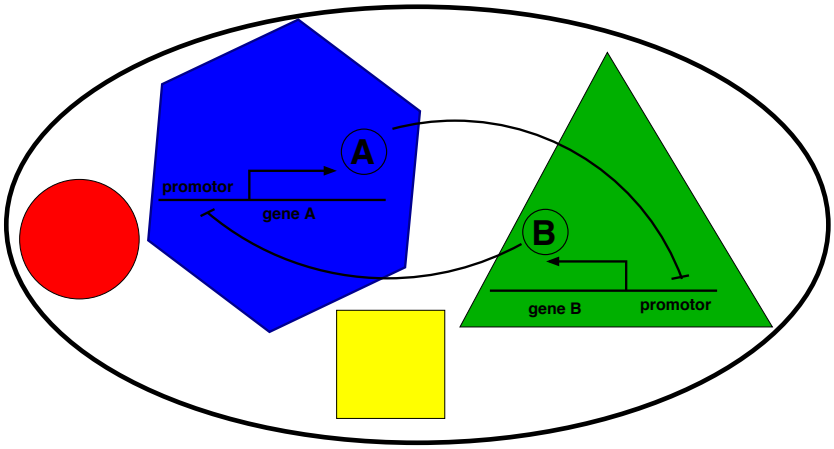

Figure 2.6.: Key regulators.
The sketch illustrates the cross-antagonism between two generic transcription factors, termed A and B, that are key regulators for the development of the "blue" and "green" lineage, respectively. Such a switch like behavior is facilitated by the ability of A to suppress transcription of B, and vice versa.

negatively regulate their mutual expression [104, 105, 106, 107], thus suppressing the competition counterpart. For the particular case of the PU.1-GATA-1 interaction, but also in the more general case, it had been demonstrated by a number of theoretical approaches that such an antagonistic action is sufficient to induce switch-like behavior [92, 108, 109, 110, 111].

2.3.3. Reversibility of lineage decisions

Within the classical paradigm of a hierarchical development from stem cell towards differentiated cell types, lineage specification is commonly perceived as series of irreversible fate decisions. This implies that once a decision is made, a cell is committed to the particular lineage and cannot alter its fate [112]. However, this restrictive view on lineage specification has been seriously challenged by a number of experimental [22, 41, 42] as well as conceptual works [113, 114]. For the example of differentiated B cells it has been reported that enforced expression of C/EBPα and C/EBPβ leads to their rapid and efficient reprogramming into macrophages [115]. Similarly, there is evidence that conversion between early committed erythroid and myeloid cells is possible if the culture conditions are changed respectively [116]. Also on the conceptual level is has been argued that a mechanism of lineage specification, which includes the potential for reversible actions, is much more flexible in response to different environmental signals [113].

With respect to lineage decisions the term *reversibility* is commonly used in

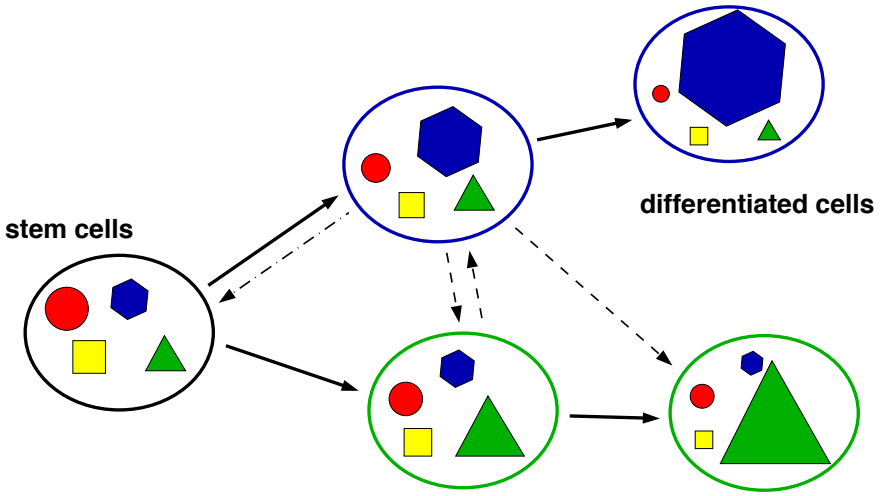

Figure 2.7.: Reversibility.
In contrast to the solid arrows indicating major developmental pathways, the concept of reversibility allows for the existence of alternative pathways. The dashed arrows illustrate the concept of fate reversibility. Such a reversion from one lineage into another might by a rare event under physiological conditions but can be achieved by targeted overexpression of certain genes. The same applies to the dash-dotted arrow indicating developmental reversibility in which the cell is shifted to a less committed state.

two, slightly different meanings: First, reversibility describes the reversion of cell fates as it is outlined for the above examples: cells can be manipulated to revert their fate to a different cell type as the one they are originally committed to. This phenomena is sometimes also referred to as "trans-differentiation". In a related context the term *plasticity* has been introduced to characterize the potential of a (e.g. hematopoietic) stem cell to revert its fate towards another (e.g. non-hematopoietic) lineage. In its second, more conceptual meaning, reversibility refers to the developmental (and molecular) reversion of a decision process by "moving the cell backwards" to an earlier developmental stage. This reversion might include the reacquisition of stem cell properties such as self-renewal ability and multi-lineage potential. Prominent example of such behavior are found in Drosophila melanogaster [117, 118] and are also proposed for the hematopoietic system [46, 119]. Such behavior is sometimes referred to as "de-differentiation".

The different aspects of reversible lineage decisions are visualized in Figure 2.7. However, it remains illusive whether a clear distinction between these two aspects is necessary and instructive. Both, trans- and de-differentiation characterize a behavior that is contrary to the major developmental pathways. The physiological and functional importance of such reversible developments remains subject to

further experimental and theoretical research.

2.3.4. Heterogeneity and lineage bias

There is increasing evidence that HSCs are an intrinsically heterogeneous population, not only with respect to their repopulation ability [45, 120, 121, 122], their cell cycle activity [66, 67] or their expression of certain surface markers [123], but also with respect to their lineage contribution. Using single cell transplantation experiments Dykstra and colleagues demonstrated that individual HSCs showed different contributions to the populations of B-, T- and myeloid cells after reestablishment of hematopoiesis in irradiated hosts [45]. In particular, the authors defined four distinct types classified according to their contribution to either myeloid or lymphoid cells. Two of these cell types with a significantly higher contribution to the myeloid lineages showed increased repopulation potential. Upon transplantation of bone marrow from the corresponding primary recipients in secondary and tertiary hosts the pattern of lineage contribution was largely reproduced, suggesting that the so called "lineage bias" is stabilized by epigenetic mechanisms. These results are in general agreement with earlier reports about the stable inheritance of the lineage bias and higher repopulation efficiency for the myeloid-biased stem cells [44]. Figure 2.8 illustrates the idea of a lineage biased, heterogenous stem cell population.

However, Dykstra and colleagues [45] also report that in the course of transplantations a slow conversion of the subpopulations towards a higher contribution of lymphoid cells and decreasing repopulation potential has been observed. In this context, it is less obvious whether a classification of a defined number of subsets is helpful for the understanding of the population inherent heterogeneity or whether a continuous spectrum does better reflect the underlying regulation.

2.4. Analysis of single cell developments

The direct experimental characterization of lineage specification in HSCs poses a number of technical difficulties. Most challenging is the analysis of the development of individual stem cell *in vivo*. Only in the case of clonal repopulation assays or by the use of retroviral markers, the lineage contribution of single cells is accessible over time. However, the location of the cells in the bone marrow and their activation patterns are subject to extensive research [71].

Alternatively, many experiments characterizing the nature of the lineage specification process in HSCs have been performed in culture systems. As outlined in the previous section, these experiments are mostly designed to capture the overall lineage contribution of a certain initial cell population without paying much attention to the temporal evolution and chronology of cellular development as it occurs within a single cell. However, it is precisely the development of each individual

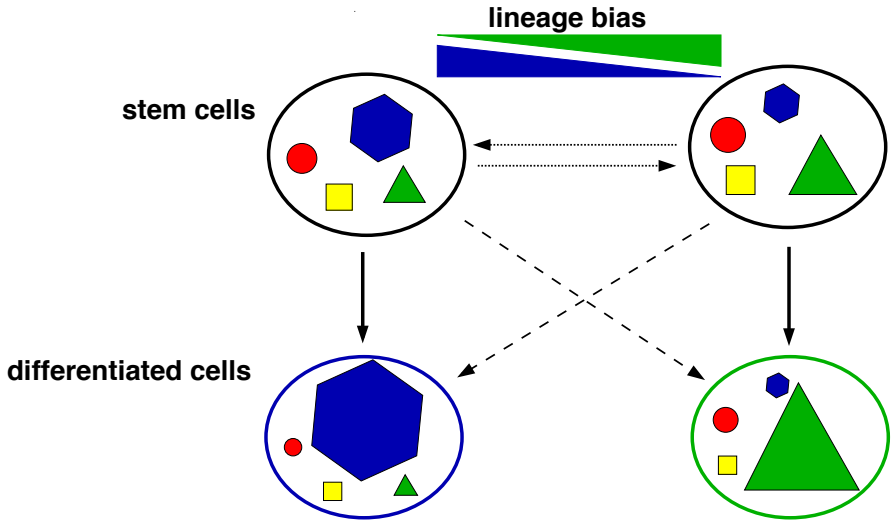

Figure 2.8.: Heterogeneous lineage contributions.
Stem cell populations (shown on top) contain a mixture of cells with varying potential to contribute to one or the other mature cell types. Such lineage biased cells primarily give rise to one particular cell type (solid arrows) and yield only minor contributions to other cell types (dashed arrows). Whether there is a slow flux between these lineage biased states (dotted arrows) is still controversial.

cell and its progeny that represents a possible realization of the developmental sequence and retains much of the necessary information: on the correlations between differentiation and cell cycle regulation, on the timing of lineage specification processes and cell death events as well as on the role of asymmetric developments [124, 125]. In this respect, it is the information about each cell's identity that allows to put the available data in its divisional context. Figure 2.9 outlines the dilemma.

It is here that the digital revolution in microscopy as well as the increasing memory capacity of computer systems opens a new dimension for the application of time lapse video microscopy for the analysis of cell cultures. Such high resolution technologies facilitate the tracing of a single cell, comprising all its progeny over extended time periods up to several days. This includes the temporal analysis of cell specific parameters like morphology, cell cycle time, motility or the occurrence of cell death within the population context. Time lapse video monitoring with single cell tracking has been applied to cultures of hematopoietic [124, 126, 127] as well as neural [128], muscle [129], and embryonic stem cells [130]. In a recent study it could be shown that the identification of patterns in the *in vitro* cell cycle time distribution proved useful for the enrichment of cells with higher repopulation potential *in vivo* [126]. Continuing these ideas, the fluo-

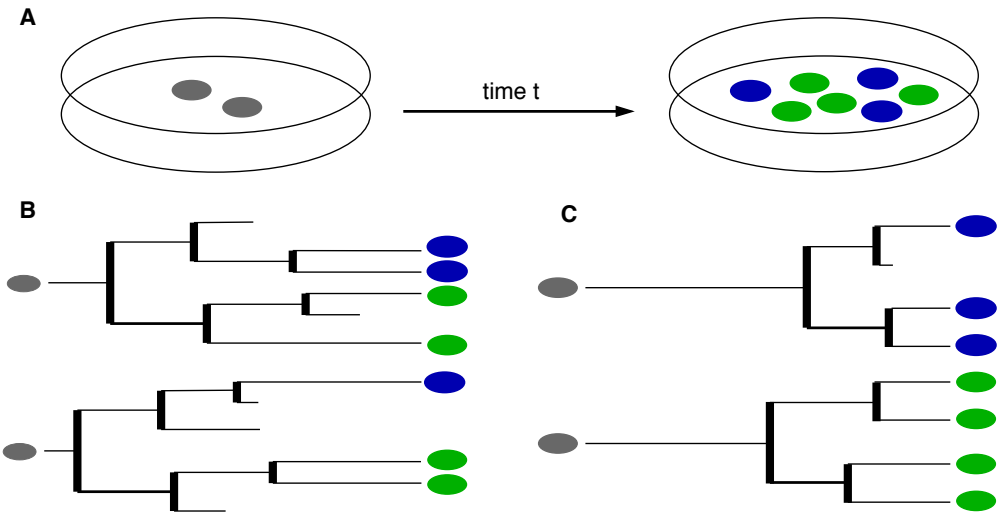

Figure 2.9.: Single cell development.
(A). Layout of a "classical" cell culture experiment for cell fate analysis in which a population of undifferentiated cells (grey) develops into a populations of cells with contributions of different lineages (blue and green). However, there is an infinite number of possibilities giving rise to this outcome. Two potential realizations are shown in **(B)** and **(C)** with different levels of division, cell death and lineage potential of the originally undifferentiated cells.

rescence labeling of marker genes for differentiation and lineage specification will soon allow a better identification and temporal determination of central decision events in the developmental sequence [131, 132]. All these different information on cellular development, divisional history, and differentiation can be comprised into a pedigree-like structure in which the founder cell represents the root and the progeny is arranged in the branches. These pedigrees represent a special class of annotated trees and are referred to as *cellular genealogies* [125].

Analyzing the time lapse video of a cell culture allows the tracking of a multitude of root cells. The resulting cellular genealogies represent unique examples of the developmental sequence as they occur under the particular assay conditions. Statistical analysis of these cellular genealogies can reveal typical patterns of cellular development as they are imprinted in the topology. Although this novel technique is pushed by a number of groups, there are no established measures for analyzing and comparing this particular type of data. However, a formal characterization of the topologies together with a set of newly developed measures is sufficient to extract the "footprints" of potential mechanisms of lineage specification that are imprinted in the cellular genealogies. A more detailed account is provided in Chapters 6 and 7.

2.5. Experimental techniques and measures

Hematopoietic stem cells are among the best studied stem cell systems for which a large body of experimental results has been accumulated over the last decades. There exists a huge number of available assays and isolation techniques to study the functionality of these cells. Some of the techniques are briefly explained as they support the motivation and understanding of this thesis.

Hematopoietic stem cell assays. As outlined in Section 2.1 the ability of a stem cell population to repopulate a depleted animal is still the major functional criteria of stem cell function. This is tested using the long-term repopulating ability (LTRA) assay which evaluates the contribution of donor cells to all hematopoietic lineages after 6 or more months [133]. Using this assay it is possible to distinguish long-term repopulating stem cells from cells that only contribute to hematopoiesis for a shorter time span (short-term repopulating cells) which do not guarantee the extended survival of the animal. Moreover, the assay also accesses the contribution to the individual lineages and is suited to detect lineage biases induced by the donor cells [45].

Alternative to the LTRA assay a number of *in vitro* assays have been developed to measure the frequency of progenitors (colony-forming unit in culture; CFU-C), stem cells (long-term culture-initiating cell; LTC-IC), or both (cobblestone area-forming cell assay; CAFC) [134, 135]. Generally, these assays are less suited to determine the lineage potential of the cell types in question.

Hematopoietic stem cell markers. Especially in the hematopoietic system, in which a prospective isolation of stem cells based on their morphology or localization is not possible, alternative methods for the purification are required. Most such techniques are based on flow-cytometry which allows to select cells based on the expression (or absence) of a set of cell surface proteins. In order to quantify these expression levels, fluorescence tagged antibodies are used which bind to the corresponding antigens on the target cells.

The most common strategy to enrich for HSCs is the selection of a subpopulation of cells that show no or very low expression for any markers of mature cell types such as TER-119 (erythroid), B220 (B cells), Mac-1 (monocytes), Gr-1 (granulocytes), or CD3, CD4 and CD8 (T cells) but high or very high expression of Sca-1 and c-Kit [136, 70]. Such cells are termed LSK (lin^-, $Sca{-}1^{\text{high}}$, $c{-}Kit^{\text{high}}$) cells. However, the purity of long-term repopulating HSCs can be enhanced using additional signals of the SLAM[1] marker family [82]. HSCs with high repopulation potential were reported to be positive for the SLAM markers CD150 but negative

[1] Signaling Lymphocyte Activation Molecule

for CD48, CD244, and CD41. Combining these markers together with the LSK markers up to 50 % of the isolated cells qualified for single cell repopulation[82].

Alternatively, dye efflux properties are regularly used to select for cell populations containing increased fractions of HSCs. For example, the exclusion of efflux of DNA binding dye Hoechst 33342 (commonly referred to as *side population*) proved useful to further enhance the purity of LSK populations [137, 70].

In the human situation, HSCs have been purified on the basis of the expression of cell surface markers such as CD34 and CD133 and the absence of glycoprotein CD38 [138]. However, in the case of CD34 expression it could be demonstrated that the expression levels depend on the cell's activation state and are subject to reversible changes [22, 23, 139].

Although the outlined purification protocols are indispensable tool to enrich for HSCs there are also a number of structural deficits connected to these methods. First, for many of the used markers it is not yet clear how their expression is functionally correlated to the stem cell function [60, 61]. Second, there is a considerable level of heterogeneity inherent to the hematopoietic system [140]. It might even be the case that a fluctuations in the expression of certain markers are ultimately required for the function of HSCs and might be reestablished after a purification procedure [61, 123].

Differentiation assays. Hematopoietic stem and progenitor cells are generally cultured in either of two conditions: self-renewing or differentiating. In the first case the culture conditions are optimized (using appropriate stromal cell lines, serum and/or combinations of cytokines) to keep the cells in a rather undifferentiated state, possibly increasing their numbers without major contributions of differentiating cells. In the second case, the culture conditions are tuned to yield a potentially pure output of differentiated cells of a special cell type. Typically, such conditions are supplemented with appropriate growth factors and cytokines (e.g. erythropoietin and stem cell factor for erythroid differentiation [141, 142] or granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) for myeloid differentiations [143]).

The differentiation status of the cells is generally assessed using morphological characteristics (supported by the use of specific staining methods), or the detection of cell type specific surface markers as there are TER-119 for erythroid cells, B220 for B cells, Mac-1 for monocytes, Gr-1 for granulocytes, or CD3, CD4 and CD8 for T cells.

Single cell development. The development of single cells and their contribution to different lineages is most commonly studied in single cell assays [26]. After plating individual cells the composition of the resulting colonies is studied using above mentioned methods.

The availability of stable transgene insertion into a host genome allows for the irrevocable and unique marking of individual cells as the random integration of the transgene allows the detection of all its descendants. Based on this method the retroviral gene marking has become an important tool for investigating the *in vivo* fate of different cell types, both in animal models and in clinical applications [144, 145, 146]. After transplantation of an appropriately treated HSC population the contribution of multiple cell clones to the different hematopoietic cell lineages can be studied over extended time periods.

Alternatively, clonal information about the development of individual cells can be provided from continuous time-lapse monitoring of *in vitro* cell cultures [147]. In contrast to the clonal tracking using transgene integration, this method allows for the derivation of complete pedigrees subtending from one initial cell. In combination with the fluorescence labeling of certain relevant marker genes for differentiation these methods hold a great potential to study the lineage specification in its divisional and developmental context.

Molecular methods in stem cell biology. Modern stem cell research benefits from the availability of biotechnological methods to study molecular processes on many different scales. Especially the establishment of the polymerase chain reaction (PCR) as a standard technique to amplify a single or few copies of a piece of DNA or RNA across several orders of magnitude has greatly supported this development. Furthermore, the development of different types of microarrays allows for the large scale interrogation of molecular content in cell populations (e.g. for the detection of mRNA transcripts (DNA microarrays), the presence or absence of proteins (protein microarray), or protein specific DNA binding sites (Chromatin immunoprecipitation on Chip; ChIP-on-chip)). Application of DNA microarray techniques in time structured experiments can elucidated temporal changes in the genetic expression profile and relate it to phenotypical and functional alterations [40, 92, 98]. The availability of even more advanced methods for the sequencing of pieces of mRNA and their simultaneous quantification using high-throughput sequencing (also termed "pyrosequencing") will dramatically influence biological research in general and stem cell research in particular.

However, among the greatest challenges with these high-throughput techniques is the statistical analysis of the resulting data, and in particular their integration into an biologically meaningful context. It is often the case that among a set of identified components and interactions just for a minority of them their role in a relevant biological processes is already resolved. For the remaining findings a comprehensive understanding of the underlying regulatory principles is missing and can hardly be predicted based on the available data alone.

2.6. Technical challenges and ethical considerations

Primary interest in stem cell research derived from the expectation to gain insight into the organizational principles of higher organisms. Beyond this academic interest it soon became clear that tissue resident stem cells hold a great promise for medical applications as these cells are in principle able to regenerate or replace diseased or mal-functioning tissue. As already illustrated in the introduction, a direct translation into medical applications is still limited as many types of tissue stem cells are rather difficult to isolate and to expand *in vitro*, are more or less restricted to a particular tissue type and, in many cases, have limited capacity to regenerate major parts of a tissue or even a whole organ.

In contrast, these limitations do not apply to embryonic stem cells. These cells are derived from inner cell mass of the blastocyst and are defined as cells that can be cultured *in vitro* for many passages and retain the principle ability to contribute to all three germ layers. However, even the availability of embryonic stem cell cultures did not overcome a fundamental problem in regenerative therapies which is the immune-incompatibility between the allogenic donor tissue and the host system, often resulting in transplant rejection.

In a remarkable experiment in 1996, Ian Wilmut and colleagues demonstrated that the transfer of the nucleus of an adult endothelial cell of a sheep into a enucleated oocyte (generally referred to a as *somatic cell nuclear transfer*) was in principle sufficient to derive a developing blastocyst [148]. After implantation of the resulting blastocyst in a surrogate mother a cloned animal (which in the first reported case became known as "Dolly", the Sheep) was established that was genetically identical to the donor animal of the original nucleus. This experiment proved that, first, all relevant genetic information is retained even in a differentiated cell and that, second, there exist a possibility to derive genetically identical ("cloned") embryonic tissues with the full developmental potential to generate a cloned animal.

Although the adaption of the experimental protocols to the human situation and the derivation of patient specific human embryonic stem cell lines is in principle possible these protocols involve a number of steps that are ethically controversial as there are the donation and utilization of human oocytes, and the manipulation of cloned human blastocysts. These techniques strongly interfere with several aspects touching the dignity of human life, and the public discussion within this area illustrates that there exists no general consensus on whether such cells should be manipulated in order to use them for the treatment of severe and potentially lethal human diseases.

However, the efforts to derive human ESC lines from cloned blastocysts declined rapidly as an alternative method for the derivation of genetically-customized stem cell lines became available. In 2006 Kazutoshi Takahashi and Shinya Yamanaka proposed the reprogramming of normal body cells into pluripotent, embryonic-like

stem cells by the use of retroviruses to transfect the cells with a set of four critical transcription factors [6]. These methods became soon available for the human situation [7, 149] and are now further refined to avoid the permanent integration of viral factors and to ensure long-term safety of the reprogrammed cells [150].

The availability of protocols for the derivation of patient specific induced pluripotent stem (iPS) cells overcomes two fundamental problems in regenerative stem cell therapy: first, these protocols do not rely on the ethically questionable usage of oocytes and the transient generation of human blastocytes and, second, the resulting cells are genetically identical to the somatic cell donor, thus avoiding complications from immune system rejection of the generated transplant. Although these protocols do not yet exist for clinical applications there is a strong research interest in this field and first results can be expected in a foreseeable future. It is the challenge of the next decade to identify experimental techniques to not only achieve the reprogramming from the somatic to the pluripotent state but to direct the differentiation from the pluripotent state towards the desired tissue type that is necessary for transplantation. At this point it is not even clear whether a complete reprogramming to the pluripotent state is ultimately necessary or whether a somatic cell can be directly reprogrammed and expanded into another, desired type of tissue cell. A fundamental understanding of the underlying processes on all organizational levels certainly requires the joint efforts of experimental biologist, clinicians and theoretical scientists alike.

3. Theoretical background: Conceptual and mathematical models

3.1. General aspects

3.1.1. Models in the natural sciences

Models are generally characterized as *representations* of a natural object or process in which not all attributes of the original are necessarily considered (*reduction and abstraction*). Focusing on the aim of a model construction and the benefit of its application, certain aspect are intentionally neglected in order to pronounce the particularly relevant features of the original and to allow the generation of acceptably accurate solutions (*pragmatism*) [151].

In natural sciences, a theoretical *model* is a construct with a set of variables and a set of logical and quantitative relationships between them to enable reasoning about the underlying processes within an idealized logical framework. Such models are an important component of scientific theories and can either be qualitative (i.e. descriptive and logical representations) or quantitative (i.e. mathematical representations). In contrast to qualitative models, quantitative models allow for an analytic, numeric, or simulation analysis [152].

Quantitative model approaches have played an important role in most natural sciences with particular success in the field of physics. However, even there it took about 300 years from the earliest works of Galilei and Kepler until revolutionary works by Maxwell, Planck and Einstein at the end of the 19th and beginning of the 20th century that finally changed the perception on theoretical sciences. Since this time theoretical physics has established as an indispensable branch of research that actively guides experimental approaches in all fields of physics.

A similar influence of theoretical research is still lacking in biological sciences. This is largely due to the nature of the studied objects which are living entities. Such objects are difficult to define since they are themselves subject to a multitude of extrinsic influences [153]. Moreover, the limited accessibility and the complexity of the constituting parts have long hampered a rigorous characterization which is a prerequisite for any quantitative model construction. Due to the increasing availability of appropriate experimental techniques on one hand and the computational resources on the other hand, quantitative approaches are now gaining more and more impact in the biological sciences. By reduction and abstraction, such theoretical models are indispensable tools to elucidate regulating principles and therefore to design experimental strategies and to verify results [154].

3.1.2. Quantitative models in stem cell biology

Although the first demonstration of a self-renewing population of hematopoietic progenitors by Till and McCulloch in the 1960s [48] was accompanied by a mathematical description of the data [155], theoretical approaches played a minor role in stem cell biology for a long time. However, there are a number of outstanding modeling contributions with significant impact on the research field as there are the works of Mackey and Glass on the dynamical behavior of hematopoietic diseases [156, 157], the works of Wichmann, Loeffler and colleagues on the regulation of granulopoiesis and erythropoiesis [158, 159, 160, 161, 162, 163] as well as on the organization of intestinal and epidermal stem cells [164, 165, 166], the works of Ogawa and coworkers on the stochasticity of cell fate decisions [167], the works of Abkowitz and colleagues on hematopoietic stem cell regulation and malignancies [168, 169, 170, 171], the conceptual works on stem cell identity by Potten and Loeffler [1] as well as their transformation in a comprehensive single-cell based modeling framework for hematopoietic stem cells by Roeder and Loeffler [33].

The above list is neither complete nor does it establish a valuation. However it demonstrates that theoretical models in stem cell biology address different biological phenomena which are also approached on different levels of description. Whereas some models are based on average responses of a cell population rather than on individual cells, others explicitly account for the variability between individual cells by modeling cellular interactions on the single-cell level. Inspired by the increasing knowledge on the molecular organization of stem cells, there is a novel class of models emerging that aims on the understanding of molecular regulation and cell-intrinsic responses to external stimuli within individual cells. However, these cell-intrinsic models have to be coupled to the cellular level to achieve an understanding of the tissue organization. The different aspects of modeling approaches in stem cell biology are subject of the subsequent Sections 3.2 and 3.3.

As outlined in the previous chapter, much of the common perception of stem cell organization derives from the idea that stem and progenitor cells can be clearly distinguished on the basis of their functional potential, their surface marker expression or even their morphology. This view is internalized in the hierarchical structuring of stem cell organization as it is shown in Figure 2.2 and is fundamentally based on the concept of distinct (stem) cell compartments. This reflects the idea that differentiation of primitive stem cells is a sequential transition through a sequence of developmental states that can by characterized by compartments with a certain set of features (e.g. expression of cell surface markers, lineage potential) and well-defined boundaries. As the stem cells pass through these subsequent compartments they progressively lose stem cell potential, restrict their lineage choices and acquire lineage specific features.

However, a functional test, like the determination of a single cell's lineage potential within a certain assay is inherently incomplete since it cannot be exclude that under different assay conditions another lineage contribution would be observed. Experiments on the independent development of paired daughter cells showed that cells immediately separated after division do not necessarily undergo the same developmental processes and contribute to identical lineages [35, 28]. This imposes a structural dilemma, since the measurement of a certain function of a cell does ultimately alter the cell itself. This phenomena has been described as the "uncertainty principle of stem cell biology" [46] and can equally be applied to experiments that aim to simultaneously measure self-renewal and differentiation capabilities or the differentiation potential along multiple lineages.

The "compartmentalization" of the hematopoietic system is closely interrelated (but not identical) to the idea that the transition through the particular compartments is always unidirectional from the most primitive cell stages down to the committed progeny. However, it has been shown that the repopulation potential and the lineage contribution of stem cells is correlated to the cell cycle [172], and that the expression of surface markers like CD34 and Sca-1 are subject to reversible changes [22, 123]. Such findings on reversibility as well as the inability to prospectively isolate *pure* stem cell populations [173] seriously challenge this concept of strictly compartmentalized stem cell hierarchy.

This controversy lead to the proposition of a concept of stem cell organization on the basis of flexible and self-organizing cellular units (agents) which are not *by definition* restricted to a unidirectional flow through a sequence of compartments (see for example [1, 34, 46, 60, 113, 174, 175, 176]). Although such adaptive systems proofed very robust in other fields of science (e.g. swarm intelligence, ant algorithms, distributed computing), there is still a level of reluctance against such novel and alternative approaches in stem cell biology.

3.2. Models of stem cell self-renewal and differentiation

It is a long-standing set of question on how the balance between self-renewal and differentiation is maintained in the hematopoietic but also in other stem cells systems, how the stem cells maintain the ability to respond to changing demands and how malignancies alter these organizational properties. In simple words, it is asked which mechanisms account for the maintenance of a stem cell population on one side and on the other side provide enough mature cells for the support and regeneration of the particular tissue. An overview over the main classes of quantitative models in hematopoiesis is provided below.

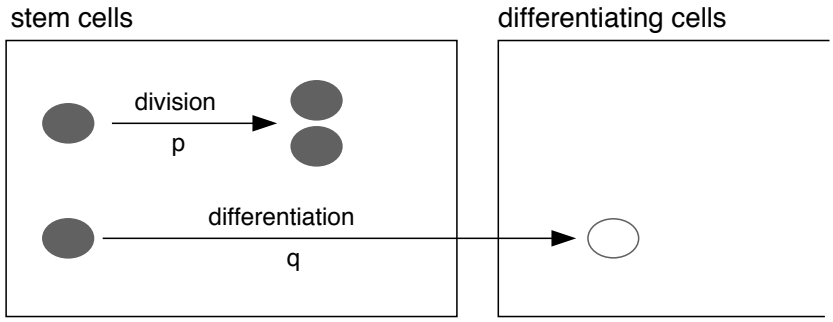

Figure 3.1.: Compartment model of hematopoietic self-maintenance.
A simple compartment model was proposed by Till and McCulloch [155]. Cells in the stem cell compartment (dark grey) are either transferred into the compartment of differentiating cells (grey circles) with rate q or they undergo division with rate p producing two daughter stem cells. In the stable situation the rates for differentiation and division are equal $p = q$.

3.2.1. Compartment models with fixed stem cell characteristics

Entity-based models. Entity-based models of stem cell organization emerged from the classical perception of stem cells as well-defined and clearly distinguishable objects. These entities are defined by a set of fixed developmental options characterizing the transition between the compartments. Such compartments are commonly envisioned to contain certain cell populations like stem cells, progenitors, or more mature cell types. The transition rules can potentially be influenced by extrinsic signals like feedback regulation from mature cell stages using growth factors or micro-environmental regulations.

A typical example of this model class is the mathematical model that accompanied the first experimental demonstration of the self-renewing ability of hematopoietic progenitor cells by Till and McCulloch in the 1960s [155]. This model assumes that within a stem cell compartment each cell can either divide symmetrically (giving rise to two stem cells which corresponds to self-maintenance of the population) or differentiate (Figure 3.1). These options are realized with fixed probabilities denoted as p and q, respectively. Even without the assumption of feedback regulation or micro-environment interaction, the authors could explain the frequency distribution of spleen colonies (CFU-S colonies) in secondary recipients.

Picking up on this idea of stochastic transitions between rather distinct cell populations, in which a probability is assigned to each developmental option, the mathematical concept was extended for the hematopoietic system [177] and for other tissues [166]. In a simple, albeit elegant approach, Abkowitz and colleagues used such an extended model to access the frequency and replication kinetics

of hematopoietic stem cells *in vivo* as well as the evolution of myeloproliferative disorders [169, 170, 171]. Besides the developmental options of symmetric division and differentiation the authors also accounted for the process of apoptosis.

Population-based models. Assuming homogeneous stem cell populations of sufficient size, entity-based systems can be described equally well by changing to population-based approaches in which the individual stochastic decision is replaced by averaged, deterministic rate constants. Feedback regulations can be introduced in both scenarios by a dynamic adaption of the either the transition probabilities that balance self-renewal and differentiation in the stochastic (single-cell based) approaches or by the corresponding rate constants in the deterministic (population-based) approaches. Additionally to the evaluation of the steady-state situation, such feedback regulations allow the analysis of the system response to various perturbations, as they occur after damage or injury. Sophisticated examples for these kind of models haven been developed by Mackey and coworkers. Under the assumptions of delayed feedbacks from later cell stages on the hematopoietic stem cell compartment the authors explained typical oscillation patterns occurring in a number of specific malignancies of the hematopoietic system [156, 157, 178, 179, 180]. Even more extensive are the multi-compartment models of later stages of hematopoiesis by Loeffler and colleagues, as there are models on granulopoiesis [159, 160, 181] and erythropoiesis [161].

Symmetric and asymmetric divisions. Till and colleagues already recognized in 1964 that the possibility of an asymmetric division (one stem cell giving rise to on stem and one differentiated cell) can be well integrated in their concept of separating division and differentiation by assuming that a symmetric division event is followed by the differentiation of just one daughter cell (Figure 3.2). This view was also supported by single cell experiments using bcl-2 transfected FDCP-mix cells [182] as well as by the conceptual works of Loeffler and Roeder [1, 46, 183]. In contrast, and driven by the observation of functionally asymmetric divisions in other stem cell system (e.g. in invertebrates [184] or in other tissues like the nervous system [185] or the epidermis [186]), the concept of asymmetric cell divisions is still strongly promoted in hematopoiesis and is present in many debates about organizational principles [94, 46, 114]. On a simplified level it is often assumed that only by means of asymmetric divisions the balance between self-maintenance and differentiation can be established. However, this balance needs to be maintained on the population level and at the present stage there is no convincing evidence for a functionally asymmetric division event in hematopoiesis at the single cell level. Moreover, the assumption of strictly asymmetric divisions of HSCs are not sufficient to explain the observed dynamic response of the hematopoietic system to injury or damage as a loss in stem cell number cannot be compensated.

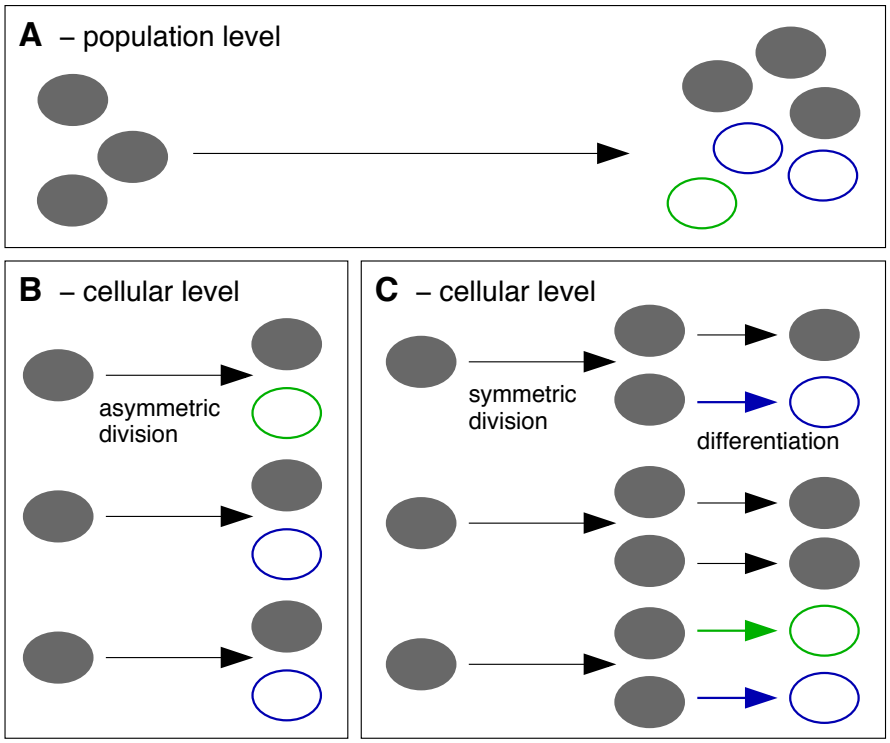

Figure 3.2.: Symmetric and asymmetric divisions.
(A). On the population level, the number of stem cells (dark grey) is maintained while at the same time a fraction of differentiating cells (colored circles) is generated. **(B).** On the cellular level this results can be obtained by assuming functionally asymmetric divisions of the stem cells in which one daughter cell remains a stem cell whereas the other daughter cell enters into differentiation. **(C).** Alternatively, the findings on the population level can be obtained if the processes of division and differentiation are decoupled. Thus, each stem cell division leads to the generation of two closely similar daughter cells (symmetric division) of which some might undergo further differentiation (colored arrows).

3.2.2. Models of adaptive and self-organizing stem cell systems

It has become increasingly evident that an entity-based view on stem cells is not appropriate to account for all facets of "stemness" including the degree of heterogeneity within stem cell populations, the reversibility of cellular developments and the adaptiveness to changing demands. Motivated by, but also stimulating these discussions, a number of stem cell models have been established that represent stem cells as individual objects with inheritable and dynamically regulated properties which only exhibit stem cell functionality in the population context. Cellular properties are continuously changing under the influence of environmental signals

and are not restricted by the boundaries of an abstract compartmentalization. For example, within such a concept the repopulation ability of a cell does not vanish at once when the cell transits from the stem cell compartment to a progenitor stage, but the the loss of function is gradual. This includes, that in a scenario with many competitor cells, a particular cell might not contribute to long term repopulation, whereas in a different situation with few competitors it could do so.

"Stemness" as a continuum parameter. The explicit conditioning on the micro-environment (either by local interaction rules or by global information and feedback loops) allows an adaptive control of the balance between self-maintenance and differentiation, generally referred to as *self-organization*. The future development of each cell depends on its current state and on the state of the surrounding environment and can only be predicted in a probabilistic sense. These single cell based models are particularly suited to account for the heterogeneity in clonal development as it occurs in chimerism studies and clonal tracking experiments (e.g. using retroviral markers). Moreover, the single cell-based structure allows to study the consequences on the tissue level that result from of each cell and can also evaluate the impact of changes of the local interaction rules on the tissue level.

Building on the early conceptual works by Potten and Loeffler [1, 62] which placed stem cell properties in the population context, the idea of "stemness" as a continuous property was formalized by Loeffler and Roeder [46]. Inspired by the special role of the hematopoietic niche for the maintenance of HSCs the model concept assumes that stem cell development is directly influenced by each cell's local micro-environment, or more generally speaking, by the signaling context to which a cell is presently exposed to. This simple assumption is outlined in Figure 3.3 in which two abstract cell properties (A and B) are modulated in a contrary fashion within two distinct signaling contexts. Subsequently, a mathematical formalization of this concept has been developed by Roeder and Loeffler [33] to construct a single cell-based model of hematopoietic stem cell organization in which the cellular development is directly regulated by the cell's micro-environment. The assumptions of reversible binding and detachment from a niche-like environment place the cells under the governance of two antagonistic regimes which is sufficient to explain self-maintenance and differentiation of HSCs as a self-organizing process. Since this model is used to accommodate the lineage specification dynamic developed within this thesis it is presented in more detail in Section 3.4 at the end of this chapter.

Another elegant approach to model stem cells as self-organizing, adaptive systems has been presented by Kirkland [34]. This work demonstrates that stem cell organization can be sufficiently described in a continuum space explicitly excluding any kind of compartmentalization in discrete subpopulations. In this concept

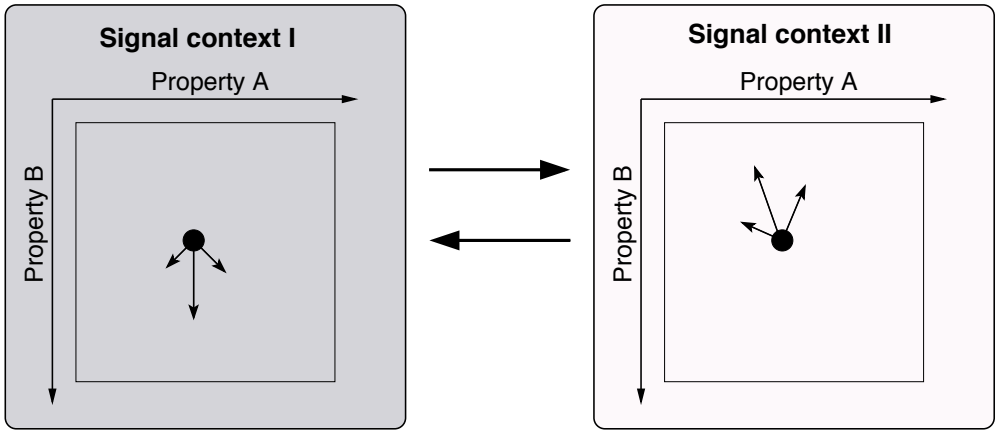

Figure 3.3.: Influence of the signaling context on cellular development. Two abstract cellular properties (named A and B) are under the governance of two distinct signaling contexts (named I and II). The signaling contexts, between which the cells can actively change, modulate the properties A and B in a contrary fashion.

cells are defined by continuous variables that describe functional and phenotypic properties (such as repopulation potential or multipotency). Based on this idea, cells can hardly be classified as "stem cells" or "non-stem cells" but can also acquire possible states in-between. The variables change according to a set of probability density functions which themselves depend on the current state of the cell: cells with higher "stemness" cycle slower and have a smaller probability for differentiation as compared to cells with reduced "stemness" (compare Figure 3.4). In this concept, self-renewal is no longer a cellular features but necessarily refers to a cell population.

These works have recently been complemented by an approach of Hoffmann, Galle and Loeffler to described a population of stem and progenitor cells as a probabilistic process that arises from cell proliferation and small fluctuations in the state of differentiation [187]. These fluctuations are assumed to reflect random transitions between different activation patterns of the underlying regulatory network and the fluctuation amplitudes are state-dependent and governed by the cells' micro-environment. The authors demonstrated that such as system reproduces the balance between maintenance of the stem cell state and terminal differentiation.

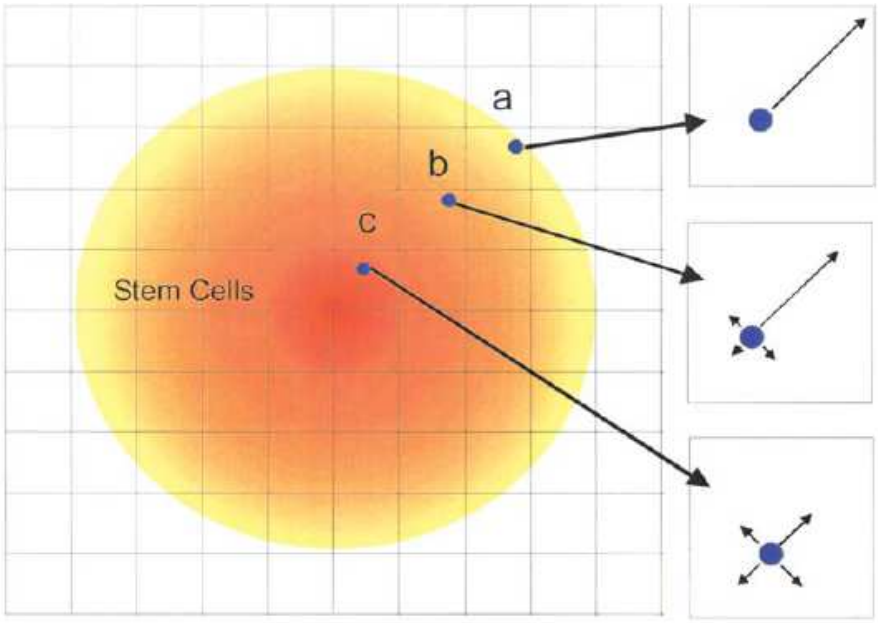

Figure 3.4.: Phase space model of hematopoiesis.
Cells at the periphery (in the light orange to yellow area e.g. cell (a)) have a high probability of moving out further towards more mature stages (white) but a low probability for "de-differentiation" towards the center. In contrast, cells close to the center (red area, e.g. cell (c)) have maximal "stemness" with slow cycling and a reduced probability of moving towards the terminal cell stages at the periphery. Cell (b) has intermediate characteristics. Picture taken from [34] with kind permission from Mark Kirkland.

Stem cells as adaptive agents. A formal foundation for the concept of perceiving cells as reactive agents, which underlies all the above mentioned models, was recently provided by Theise and d'Inverno [174]. The authors argue that critical phenomena of the system only emerge if a sufficient number of reactive agents are present. Moreover, self-organization of a complex adaptive system requires interactions between the agents as well as a certain degree of non-determinism. The authors demonstrate that the perception of stem cells as reactive agents is a possible tool to relate the phenomena on the single cell level to the resulting population dynamics of stem cells. In this sense the characteristics defining "stemness" on the population level can possibly be translated to a set of rules on the single cell level.

3.3. Conceptual approaches on lineage specification

The mechanisms that regulate the maintenance of multipotency are closely related to the more general aspects of self-maintenance of a stem cell population as outlined above. However, also in a historical context, the conceptual approaches towards lineage specification have been discussed rather isolated from the question of stem cell maintenance. Therefore, the following general overview about concepts of lineage specification and their mathematical implementations is completed by a brief discussion on how these models integrate with the conceptual ideas about stem cell self-maintenance.

On a rather simplistic level it is apparent that the transition from a multipotent cell stage towards a functionally fixed cell type is ultimately coupled to a decision process which favors one lineage at the expense of the others. The question remains how this process of lineage specification is organized on the molecular level and how it can be influenced by external queues. Furthermore, it is also not clear how much flexibility is associated with the processes of lineage decisions. This in turn, refers back to the discussion about reversibility of general stem cell characteristics in Section 2.1 as well as in the above sections.

3.3.1. The concept of developmental landscapes

Due to the complexity of intracellular regulation in lineage specification the number of analytical approaches is limited. However, it is long recognized that processes of lineage specification in adult stem cell systems have a close counterpart in embryology where a zygote gives rise to a complete organisms composed of a multitude of different tissues. In 1956, Conrad Waddington pictured this decision sequence by a ball rolling down an incline with different branching ridges and valleys [188, 189] (Figure 3.5). Whereas the valleys correspond to the different developmental options the ridges represent the distinctions between the tissue types. This neat visualization of the process of lineage specification in development received a comeback in the last years [190] and is now commonly used especially in the discussions about reprogramming and induced pluripotency [191, 192, 193].

Although the building principles of a lineage specification process are already well captured by this early visualization of a developmental landscape, the picture deserves a critical interpretation and discussion. Starting from the principle assumption that the phenotype of a cell is the direct result of the activity of the cell's genes and gene products it follows that lineage specification (i.e. the change in the cell's phenotype) is ultimately linked to a change in the underlying molecular expression pattern. The question remains, which molecular components interact with each other, how these interactions are characterized and how a robust and self-stabilizing expression pattern of these components appears.

A larger number of theoretical works on the dynamics of such large interacting

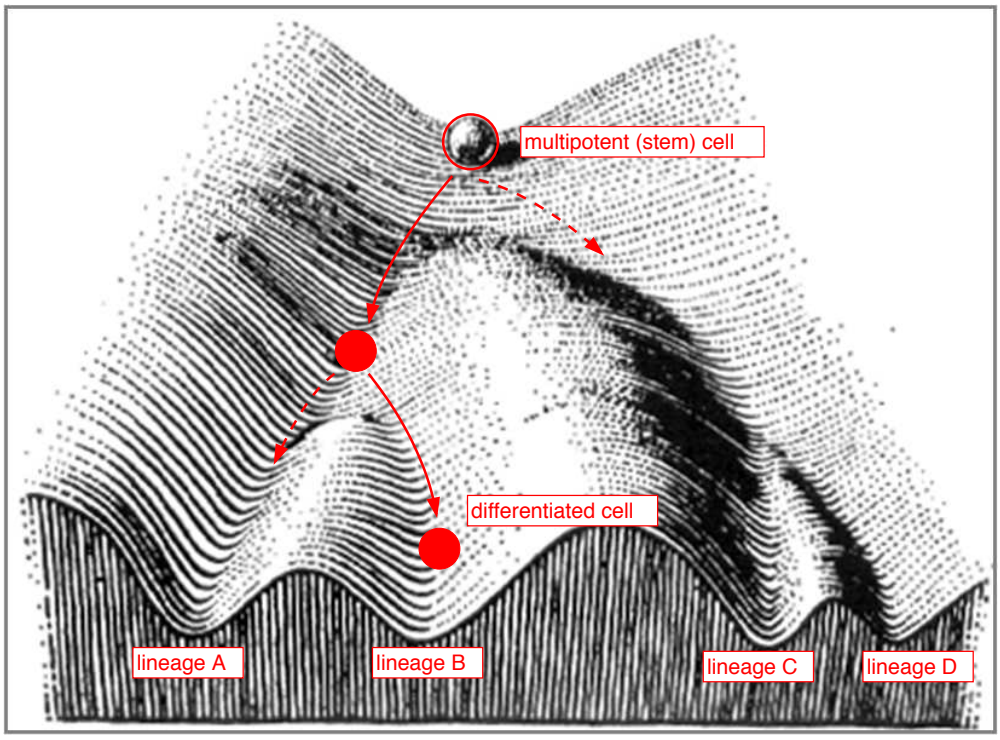

Figure 3.5.: Lineage specification in the concept of state space.
The developmental landscape as suggested by Waddington [188, 189] provides the background to understand lineage specification as a decision process in a high dimensional landscape that arises as the consequence of complex intracellular interaction network. A possible development is shown by the red arrows, alternative options are indicated by the dashed arrows. Figure modified from [189].

systems are in the focus of current research for the conceptual understanding of lineage specification. It is already evident that the valleys and ridges proposed by Waddington [188] (compare Figure 3.5) do not correspond to physical structures or potential energy landscapes as they are known from physics but rather represent unstable and stable states of the cell's underlying complex interaction network [190, 194]. Such landscapes describing the probability for finding a cell in a certain state can already be computed for smaller systems [92, 195] and provide very helpful and intuitive interpretations of the underlying network dynamics.

3.3.2. The concept of state space

It should be pointed out explicitly that the network dynamics (i.e. changes in gene expression and the existence of steady states) are determined by the particular arrangement and strength of the network interactions. Assume for example that

gene A represses gene B and that genes C and D are mutual activators. Given that these interactions are fixed based on the promoter structure of the genes, only certain expression patterns are possible while others are not. For example, A and B cannot both be activated simultaneously if one inhibits the other; conversely, C and D must both turn on if they are mutual activators. Interactions like these limit the number of self-adapting expression states and determine how they evolve over time. Given a certain stimulus the internal state of a cell might transiently shift through various intermediate stages before it finally stabilizes in one of the self-adapting expression states [196].

In a more formal description, every cellular component (gene transcripts, proteins, soluble factors, etc.) within a certain cell can be represented as a dimension in a high-dimensional *state space*. In such a state space each cell can be identified by a vector $S = \{S_1, S_2, \ldots, S_n\}$ in which each component S_i represents the concentration of a particular molecular component X_i. Given the general consensus that the phenotypic appearance of a cell is the result of the molecular interactions between all the genes, their transcripts, proteins, signaling pathways and other intracellular components, this also implies that the dimensions of the state space are not independent of each other. Such a complex, high-dimensional system of interactions is most conveniently represented by a complex network in which the nodes correspond to the different molecular components X_i whereas the edges characterize their mutual interactions x_{ij}. It is now the supportive and repressive nature of these mutual interactions x_{ij} leading to the emergence of patterns in which certain components X_i are expressed while others are silenced. The ability to maintain such an internal equilibrium is termed *homeostasis* whereas the stabilized, self-adapting patterns are referred to as *attractors* of the network. On the phenotypic level such a stable attractor can be associated with a certain cell type (e.g. stem cells, mature cell types) being characterized by a certain characteristic expression of molecular components and fulfilling particular functional tasks [197, 198].

In the above interpretation, the picture by Waddington is a representation of the intracellular attractor landscape and the developmental options between them. However, there is a central, although fundamental question which is not captured by the sketch: Is this attractor landscape fixed or is it changing in the course of differentiation? In terms of the state space formulation the cell's intrinsic state is characterized by the vector of concentrations S_i for each of the intracellular components. However, it is the the matrix of all potential interactions x_{ij} that actually shapes the attractor landscape and that determines how the cell state S_i develops within the state space. As the concept of the developmental landscape is a simplification of a highly complex process, there is no unique answer on whether this landscape is constant in time or not. In fact, the answer depends on the assumptions and interpretations made for the underlying molecular network. Two contrasting perspectives are outlined below:

- **The local view.** In the *local view* the network contains only a limited number of important molecular components for which the dynamic behavior is subject to a quantitative analysis. The concentrations of other components that are constant over time or that only change on long time scales are not explicitly described as nodes of the network but are represented as (constant) parameters. However, such parameters characterize the nature and intensity of the interaction between two or more network components. As an example one might think of transcription rates that can be described as functions of DNA methylation patterns. If such patterns change on long time scales also the depending transcription rates change slowly over time. As a consequence of this interpretation quantitative (and qualitative) changes of the network interaction ultimately alter the network dynamics and potentially modify the attractor landscape.

  As for the example of bistable switches, which is illustrated in more detail below, it is the modification of a bifurcation parameter, commonly a generalized transcription rate, which moves the system from a state with just one attractor towards a state with multiple such attractors [110, 108]. Furthermore, it could be shown by studying boolean dynamics on larger networks that modifications of the interaction rules can (although they do not necessarily have to) lead to dramatic effects on the attractor landscape [199]. On the basis of these findings, the processes of differentiation and lineage specification, which are evidently coupled to specific epigenetic changes and are closely regulated from external signals, can be interpreted as alterations of the developmental landscape.

- **The global view.** In contrast to the *local view* the *global view* represents the concentrations of each and every relevant molecular component (chromosomes and genes, transcripts, proteins, receptors etc.) within the underlying network structure. In this framework all the parameters that influence the system behavior in the local view are internalized as additional nodes of the network graph. In principle this approach can be applied to all external influences as well (like concentrations of external cytokines and nutrition, temperature etc.) leading to a significantly increased dimension of the state space vector S.

  It is the idea of the global view that interactions between any of the network components are solely characterized by physical and chemical properties and do not involve any temporal dependence. As all external influences and changes that occur even on slow time scales (e.g. epigenetic remodeling) are already internalized within the network structure, the interaction rates represent absolute and conserved properties. This, in turn, leads to a fixed developmental landscape in which all possible attractors are already

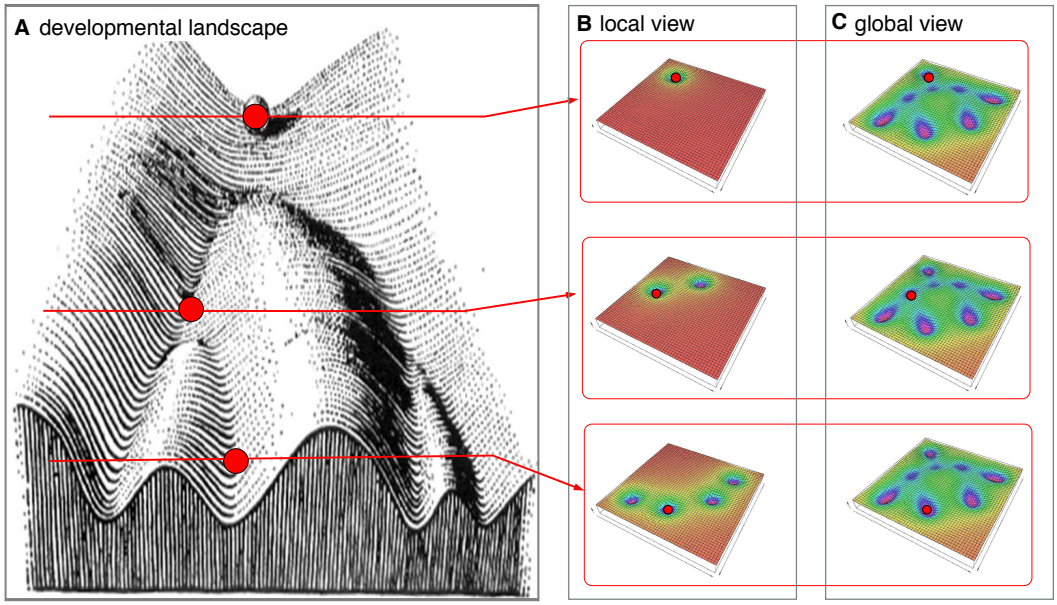

Figure 3.6.: Interpretations of the developmental landscape.
The developmental landscape in **(A)** are interpreted in the *local view* **(B)** and in the *global view* **(C)**. For three different time points a possible, two-dimensional landscape model for the state space is provided. The actual cell state is represented by the red dot.

represented. However, as the number of interacting components is large, the resulting landscape is diverse and the attractors are possibly scattered over the whole state space. This view closely resembles the rigorous definition of the a cellular state space first formulated by Kauffman [197, 200].

Coming back to the developmental landscape as it has been proposed by Waddington, this picture can be interpreted in two different ways, outlined in Figure 3.6. In the *local view* the attractor landscape changes as the interaction parameters are functions of the time of differentiation (depicted as a sequence of pictures in Figure 3.6(B)). However, it is a modulation of the landscape which drives the cell state (depicted in red) into another (evolving) attractor while previous options vanish. In this interpretation, the developmental landscape shown in Figure 3.6(A) has to be interpreted as a time-ordered sequence of snapshots rather than a fixed landscape.

In contrast, in the *global view* the landscape remains fixed (Figure 3.6(C)) and only the trajectory of the cell in the simplified, two-dimensional state space is a function of time. In this sense, there is a coexistence of multiple attractors corresponding to the multipotent and the subsequently restricted developmental

states, and differentiation is illustrated as a transition from one attractor towards the next attractor potentially involving the decision between more than one option. It should be clearly stated that the low-dimensional representation in Figure 3.6(C) is insufficient to captures the complex, high-dimensional state space in the global view but it illustrates the coexistence of several local attractors in a static landscape.

Taking together it needs to be emphasized that the two different views on developmental landscapes are not mutually exclusive. In fact, the local view appears as a low-dimensional projection of the global view. Therein some dimensions of the global view (i.e. concentrations of certain molecular components) are collapsed into interaction parameters of the network such as transcription and inhibition rates. The preference for one or the other view should depend on the particular problem in question.

3.3.3. Fluctuations and robustness

Even within a stable attractor of the state space the concentrations of the intracellular components S_i might fluctuate around the optimal state. Such fluctuations occur naturally and are the results of small copy numbers, degradation and diffusion processes within the cells (intrinsic noise) or spatial arrangements and influences of the local micro-environment (extrinsic noise) [201, 202, 203]. Noise effects can be further enhanced in non-linear systems [204].

However one needs to distinguish whether the "shape" of the attractor and the noise amplitude are suited to induce a change of the attractor state or not. Generally, small fluctuations do in most cases not lead to a change of available attractors and thus do not significantly change the overall expression pattern and the cells phenotype [205]. This feature is commonly referred to as *robustness* and is illustrated by the width of the attractors in Figure 3.6.

In contrast, and most likely induced by extrinsic effects, fluctuations of sufficient amplitude can also contribute to phenotypic variability. In fact noise induced fluctuations are a possible stimulus to induce changes of the cell state and to shift the cell into a different attractor (i.e. into another phenotype) [206, 207, 208, 209].

In the concept of developmental landscapes as it is illustrated in Figure 3.5, the fluctuations do in principle enable the cells to move from one valley (i.e. attractor) into a separate valley. The probability for such state changes depends predominantly on the noise amplitude and on the "location" of the alternative attractor. As outlined above, in the global view the picture of a low-dimensional developmental landscape is oversimplifying as the alternative attractors might be distant and separated through a rugged developmental landscape. In the local view this translates into the question of whether a certain alternative attractor coexists for a given parameter configuration or whether it needs to be recreated by a changing parameter configuration e.g. encoding epigenetic remodeling.

3.3.4. Reduction of complexity

The complexity of the state space representation of a cell including the characterization of all the mutual interactions between the cellular components makes a comprehensive approach to the cell intrinsic dynamic in the *global view* numerically very challenging, in fact not feasible at the moment. However, it is not just the feasibility, but also the availability of the particular data which limits a global approach. Although the availability of high-throughput measures like mRNA-microarrays allow to simultaneously determine the expression of thousands of genes in a cell population, these methods are rarely adequate to characterize the complex interactions between them. In the context of the cell's state space, the current state S_i can be roughly determined by these methods but the interactions x_{ij} remain unknown. Consequently a dynamic picture of the cells behavior in a global state space can not be obtained.

In order to reduce this level of complexity for descriptive approaches, the whole phenomena of intracellular regulation and lineage specification is more successfully addressed in the *local view*. Here, one may roughly distinguish between two principle approaches: a *phenomenological* approach and a *functional* approach. Figure 3.7 outlines these principle perspectives. Whereas the cell on top is depicted with different sets of competing and potentially interacting genes and gene products as well as their mutual interactions (*global view*), the snapshots below refer to the low-dimensional projections (*local view*). The phenomenological approach summarizes the complex regulations within coregulated gene sets into different major components and studies the interactions between them. Although this approach misses the molecular details of the interactions within the components, it is sufficient to address the general principles of developmental phenomena like lineage specification and the temporal ordering of events. In contrast, the functional approach focuses on the interactions between certain key regulators, for which details of the molecular regulation are available. Although this process can elucidate the role of molecular switches for certain intracellular decisions, it captures just a brief sequence out of the complex and temporally extended differentiation dynamic and fails to embed these actions into a greater context. For both the phenomenological and the functional approach a brief overview about selected models is provided below.

3.3.5. Phenomenological models of lineage specification

Driven by the experimental findings on the separate differentiation of paired progenitors using HSC [35, 27] Ogawa and Mosmann developed an early, mechanistic approach to describe lineage specification [167]. From the observation that the daughter cells gave rise to different colonies with varying contributions of hematopoietic lineages, the authors concluded that the lineage potential is pro-

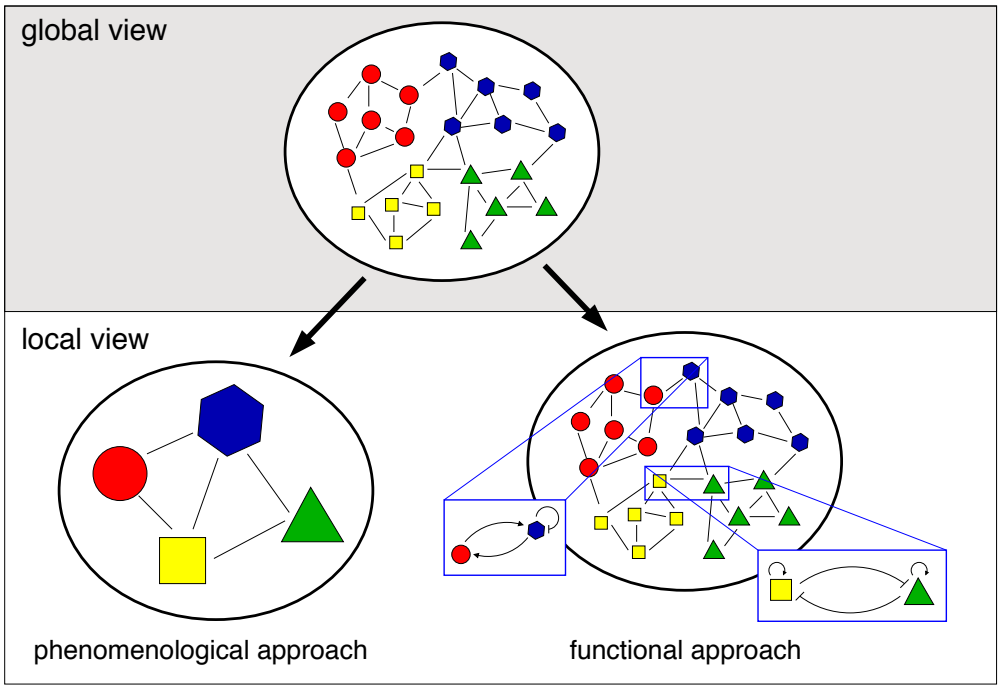

Figure 3.7.: Levels of description.
The cell on top reflects the entire state space of a cell in the *global view* (grey shaded area) with a multitude of interacting molecular components (e.g. genes, transcripts, proteins, receptors etc.). The representations below illustrate different variants of the *local view*. The left arm outlines the *phenomenological* approach in which complex regulations within coregulated gene sets (marked with identical colors) are summarized into different major components. In contrast, the *functional* approaches (shown for the right cell) focus on interactions between certain key regulators for which details of the molecular interactions are known.

gressively lost during differentiation. In particular, the authors assumed that a multipotent progenitor cell undergoes multiple cell divisions in which the daughter cells pass on a subset of the parents lineage potential. The decision of whether the lineage potential had been restricted to a single lineage, a certain number of lineages or the full potential of the parent, is characterized as stochastic (see Figure 3.8). Although the model could conceptually account for the experimental observations, it does not address the question how the stochasticity of the decision process is intrinsically represented within the cell. Moreover, it is not convincingly demonstrated why the process of lineage restriction is stringently coupled to the decision process itself. It could well be the case that lineage restriction occurs during the subsequent development of the daughter cells.

It is the particular merit of Stuart Kauffman [197, 200] who established the

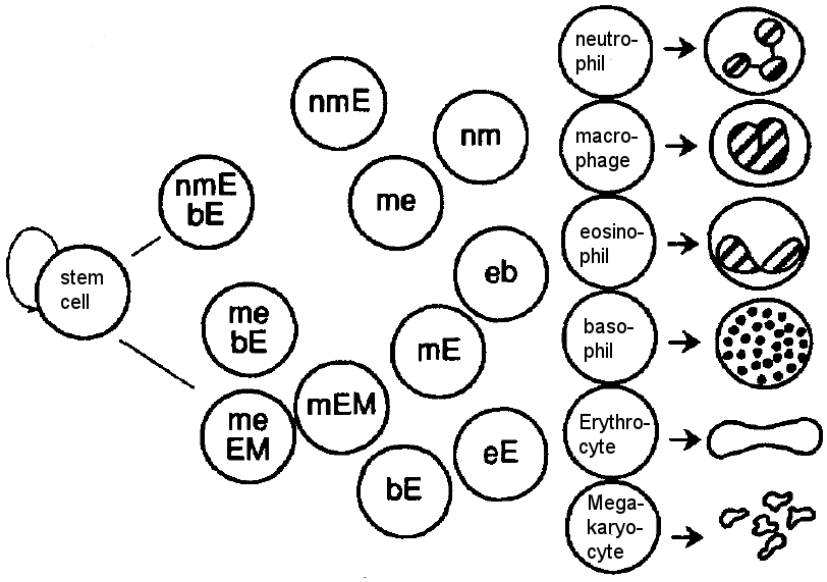

Figure 3.8.: Progressive restriction of lineage potential.
The sketch (adapted from [210]) illustrates the model first proposed by Ogawa and Mosmann to describe lineage potential as a progressive loss of lineage potential [167]. Starting from a multipotent stem cell shown on the left side, upon each division the number of possible developmental options (indicated by the letters within each cell) is randomly reduced until the daughter cell is finally committed to a certain lineage (shown on the right side). The developmental options are encoded as follows: n, neutrophil; m, macrophage; e, eosinophil; b, basophil; E, erythrocyte; M, megakaryocyte.

theory of dynamically and self-regulating networks in the realm of modern molecular biology. Using simplified, boolean network dynamics he and others [196, 199] showed, that even highly complex networks finally converge towards a finite number of possible attractors. Typical phenomenological features of differentiating cells could be recaptured by studying the number of possible attractors, their basins of attraction and the robustness of the attractor states against perturbations. Attributing different cellular phenotypes (e.g. undifferentiated or lineage specific cell states) to the different attractors, the transition between them corresponds to the process of differentiation. As Huang pointed out [205], the transition to another attractor requires a perturbation possibly involving multiple genes. Such a perturbation might be initiated via a external regulatory stimuli that trigger receptor mediated signals and often target a set of key regulators. The set of affected genes and proteins that finally change their expression is defined by the interaction network and might be the evolutionary result of the cell's response to appropriate environmental signals [205].

In a sequence of papers Kaneko and coworkers studied the parallel existence of multiple stable attractors using medium sized networks described in terms of ordinary differential equations [211, 212, 213, 214]. Their modeling approaches are based on the idea that each cell is characterized by the concentrations of a set of chemicals which participate in a network of autocatalytic biochemical reactions. Furthermore the internal chemical concentrations are influenced by the gradient of chemical concentration within the cell, its surrounding medium as well as neighboring cells. Kaneko et al. showed by a temporal evolution of the internal cell states that the final attractors for each cell are highly depending on the initial conditions and on the nature of the cell-environment interaction.

It needs to be mentioned that non of the above outlined approaches could successfully integrate direct biological parameters into the particular quantitative models. In spite of this, the particular class of models gives reason to assume that stable attractors are a common feature of large dynamical networks as they are observed for gene and protein interactions within a cell and that they are appropriate representations of a cell's internal state. In this respect the models provide a phenomenological understanding about the connection between the existence of stable expression patterns in complex networks and their translation into cellular phenotypes. The analytical studies on the basins of attraction and on the robustness of the attractors provide insight into the building principles of phenotypic heterogeneity and the stability of the phenotypes against minor and major perturbations.

3.3.6. Functional models of lineage specification

In contrast to the phenomenological approaches there is another class of models focusing on the functional understanding of the mutual interactions between key regulators. Such interactions are typically studied for small networks with a focus on rather well characterized components and their interaction dynamics.

It was not until the 1990s that the advances in the characterization of molecular interactions led to the identification of certain key regulatory elements which in turn opened the gate for mathematical modeling approaches of specific biological systems and established the concept of molecular switches related to differentiation and lineage specification. It has been shown that simple regulatory genetic networks (intrinsically coupled by positive or negative interactions) can exhibit different, separable stable attractors which are closely linked to different developmental options. The ability to change between these different stable states is the defining criteria of a molecular switch. As a demonstrative example, Gardner and colleagues presented a synthetic genetic toggle-switch in E. Coli [110] (Figure 3.9). The switch, which is constructed from two repressible promoters in a mutually inhibitory network and for which the authors also provide a simple mathematical model, could be flipped by transient chemical or thermal induction.

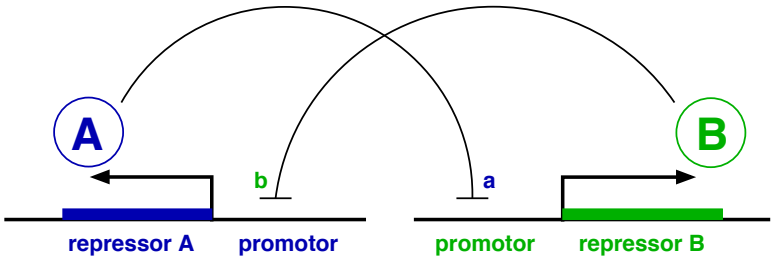

Figure 3.9.: Network with two mutual repressors.
The sketch illustrates a simple genetic network of two mutually repressive genes, similar to the model for the toggle switch proposed by Gardner et al. [110]. Depending on the cooperativity of the interaction and strength of the interaction parameters a and b the system can change from a monostable to a bistable regime.

The authors proofed the existence of parameter configurations in which just one stable attractor exists (monostability) or, for changing parameters, in which two such attractors exist (bistability).

With particular focus on hematopoiesis, the decisions of myeloid progenitors to differentiate either into erythroid/megakaryocytic cells or into myeloid cells (reviewed in [99, 215]) has been subject to different modeling approaches. It is the mutual regulation between zinc finger factor GATA-1 and the Ets-family transcription factor PU.1 that supports the idea of a molecular switch. A representative example has been proposed by Roeder and Glauche [108] and is briefly illustrated in Figure 3.10. A detailed analysis of the possible molecular interactions between the two transcription factors revealed a rich manifold of possible dynamical phenomena. Whereas in the case of low transcriptional activity the system is characterized by a stable state in which both transcription factors are equally expressed, increasing the transcriptional activity of the autoregulation of GATA-1 and PU.1 leads to the emergence of different stables states that are characterized by the dominance of either one of the two factors, thus revealing the typical properties of a molecular switch. A sequence of corresponding attractor landscapes is shown in Figure 3.11 illustrating this qualitative change in the number of attractors. In the bistable regime, the system is fixed in either of the two attractors and a switch between these attractors requires an external perturbation. In this interpretation, the bistable switch can be mapped onto the developmental landscape outlined in Figure 3.5 in which the different attractors are separated by the ridges in-between.

Supported by experimental results this PU.1-GATA-1 switch has been further studied by Laslo et al. [109] and Soneji et al. [140]. Related switch dynamics have been suggested for the differentiation of T-lymphocytes based on the interaction of transcription factors GATA-3 and T-bet [216, 217]. In addition to the works on

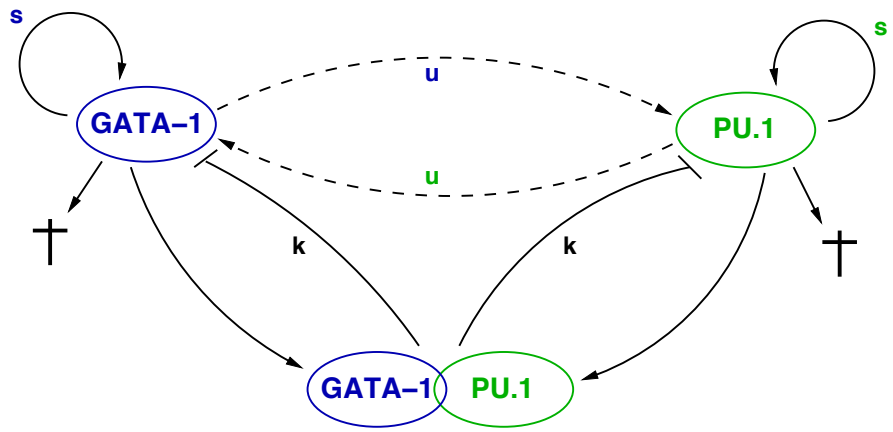

Figure 3.10.: Interaction network for transcription factors GATA-1 and PU.1.
The sketch shows a possible configuration of the interaction network between GATA-1 and PU.1 as discussed in [108] for a completely symmetric system with following parameters: s - transcription rate of the autoregulation, k - inhibition of transcription by the heterodimer, u - mutual activation (this process is most likely indirect). Crosses indicate decay. As shown in [108] this interaction network can be translated in a set of dimensionless ordinary differential equations.

bistable switches, Cinquin and Demongeot [218] and Foster and colleagues [219] proposed more generalized, higher dimensional switches that facilitates decision events between more than just two options.

Although the molecular description of intracellular interactions is most likely the best way to finally achieve a *global view* on lineage specification in which all interacting components are represented, the limited knowledge about the nature of interactions but also the technical challenges to describe such highly complex systems in an appropriately mathematical framework make this vision not yet feasible. Even for the small-scale networks discussed above, many of the interaction parameters are not accessible. However, taking the *local view* is not necessarily a disadvantage but offers the ability to study complex systems in a bottom-up approach, thus getting a notion about the functional role of different network motifs.

For the example of the discussed molecular switches it is demonstrated that external influences or changes in the general accessibility of the genes are not encoded as additional network components but are attributed to the interaction parameters (e.g. the transcription and inhibition rates s, u, k in Figure 3.10). In this sense, the changes of these secondary parameters modify the attractor landscape and induce switches between different stability regimes (e.g. mono- vs. bistability). Such a behavior is a typical consequences of the *local view*.

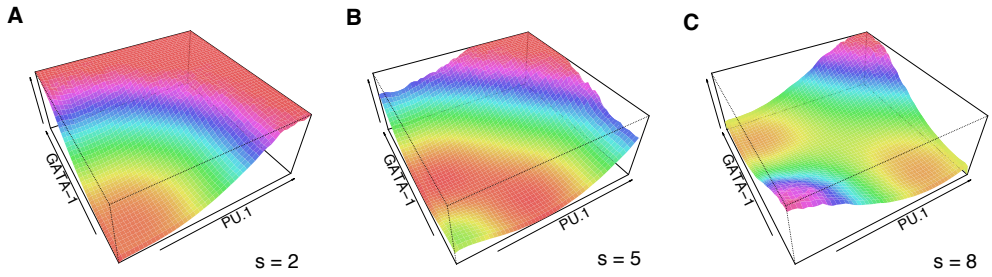

Figure 3.11.: Attractor landscape.
(A - C) show a sequence of attractor landscapes for different parameters of the transcription rates s for the PU.1 - GATA-1 model introduced by Roeder and Glauche [108]. The z-axis corresponds to the negative logarithm of the probability for finding a cell at a certain position in the state space of PU.1 and GATA-1 concentrations. Thus, the attractor regions are indicated by the orange color.

3.3.7. Integration of lineage specification and self-maintenance

Although the phenomenological models of lineage specification fail to substantiate the network topologies by experimental data, and the molecular networks are limited to rather specific components, the whole body of evidence provides an insight in how the process of lineage specification is potentially represented at the molecular level. However, it is a defining feature of stem cell populations to remain in an uncommitted state from which the lineage specification towards different cell fates is still possible. This state of maintenance of multipotency is closely correlated with the maintenance of the stem cell population itself. This interrelation between multipotency and stem cell self-maintenance is not reflected by any of the above model approaches.

In this context the question arises, how the attributes of a stem cell population as being adaptive to changing demands and the options of flexibility and reversibility extend on the process of lineage specification. Furthermore, it is unclear how the maintenance of the experimentally observed priming behavior (to low-level co-expression of lineage specific, potentially antagonistic genes) is efficiently coupled to the maintenance of the stem cell population and the corresponding environmental cues.

Based on the idea that lineage specification is a competitive interaction between different lineage specific programs a comprehensive concept is developed in the next chapter integrating the ability for lineage specification and the self-maintenance of a stem cell population into a common conceptual framework.

3.4. Roeder and Loeffler model of hematopoietic stem cell organization

The single cell based model by Roeder and Loeffler [33] develops from the idea, that stem cells have the ability to independently respond to a multitude of environmental signals. The cooperation of many such cells results in the overall appearance of the tissue specific stem cell system.

The special role of the hematopoietic niche for the maintenance of HSCs has already been outlined in Section 2.2. Building up on the idea that different configurations of the cell's local micro-environment propagate different signals and developmental fates (see Figure 3.3) a quantitative model for the HSC system has been developed. In this model the cells are in principle exposed to two contrary signaling contexts. The first signaling context, termed $\mathcal{A}$, is commonly associated with the hematopoietic niche and supports cellular quiescence and the maintenance of the repopulation potential. In contrast, the second signaling context, termed Ω, is associated with the decoupling from the niche and supports proliferation and loss of repopulation ability which finally leads to differentiation.

The model by Roeder and Loeffler [33, 220] describes each cell as an individual object with a discrete set of characteristic variables which are updated for all cells at discrete time steps typically measuring one hour. The actual status of a stem cell (i.e. its position in the state space of the defining variables) is characterized by its current signaling context ($\mathcal{A}$ or Ω), its position in the cell cycle c (indicating the cell is in either G1, S, G2, M or G0 phase) and its affinity a which quantifies the propensity of a particular cell to reside in signaling context $\mathcal{A}$. It is this affinity a which is itself modulated by the action of the two signaling contexts and which directly influences the transition probabilities. Whereas the non-proliferative cells in $\mathcal{A}$ are maintaining or even regaining their affinity a up to an upper limit $a_{\max} = 1$, the proliferative cells in signaling context Ω gradually lose their affinity a. It can be shown that the affinity a is a good indicator for the "stem cell quality" as it approximates the cell's ability to realize long-term system repopulation. An overview of the model setup is provided in Figure 3.12.

Accounting for the presumed underlying complexity, transitions between the two signaling contexts are described by a stochastic process. The probability of switching depends on the actual value of a as well as on the number of cells in the target signaling context. The transition probabilities are described by the following sigmoid functions:

transition from $\mathcal{A}$ to Ω:

$$\omega(a, N_\Omega) = \frac{a_{\min}}{a} \cdot f_\omega(N_\Omega) \tag{3.1}$$

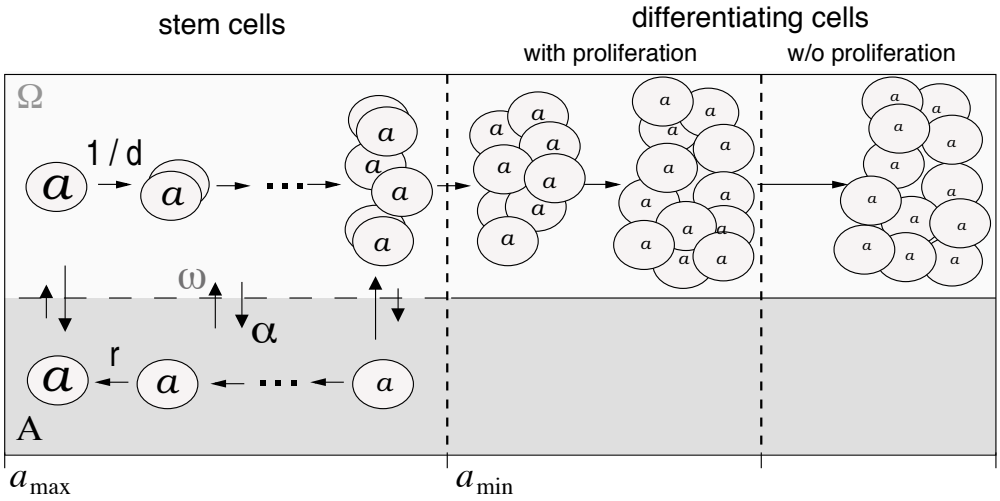

Figure 3.12.: Model of HSC organization.
The model proposed by Roeder and Loeffler [33] is characterized by two different signal contexts ($\mathcal{A}$ and Ω). Cells can reversibly change between $\mathcal{A}$ and Ω with probabilities α and ω which depend on the cell numbers and the cell specific affinity a (font sizes indicate the value of a for each cell). Whereas activated cells in Ω undergo divisions and exponentially degrade their cell specific affinity a with rate d, cells in $\mathcal{A}$ are quiescent and preserve/regain their affinity a with rate r. Cells with affinity $a < a_{\min}$ cannot change into $\mathcal{A}$ again and undergo final differentiation (with and without proliferation).

and transition from Ω to $\mathcal{A}$:

$$\alpha(a, N_A) = \frac{a}{a_{\max}} \cdot f_\alpha(N_A) \tag{3.2}$$

Within these equations N_Ω and N_A refer to the cell numbers in signaling context Ω and $\mathcal{A}$, respectively. The parameter $a_{\min}$ is an arbitrary, but numerically necessary boundary to account for the fact that cells with $a < a_{\min}$ have a significantly reduced probability for transition into $\mathcal{A}$ but a high probability for transition into Ω. In other words, cells with $a < a_{\min}$ do hardly account for the maintenance of repopulation ability within $\mathcal{A}$ and are thus referred to as differentiating cells.

The functions $f_\omega(N_\Omega)$ and $f_\alpha(N_A)$ are sigmoid functions of the type:

$$f(N) = \frac{1}{\nu_1 + \nu_2 \cdot exp\left(\nu_3 \cdot \frac{N}{\tilde{N}}\right)} + \nu_4. \tag{3.3}$$

Whereas the parameters ν_1, ν_2, ν_3, and ν_4 determine the shape of f, the parameter $\tilde{N}$ is a scaling factor for N. It is possible to transform the shape factors

ν_1, ν_2, ν_3, and ν_4 into the more intuitive values $f(0)$, $f(\frac{\tilde{N}}{2})$, $f(\tilde{N})$, and $f(\infty) := \lim_{N\to\infty} f(N)$. Details are provided in [47].

If a cell does not transit between the signaling contexts (with probabilities $(1-\omega)$ and $(1-\alpha)$, respectively) the cell's affinity a as well as the internal cell cycle clock c are modified according to the following rules. For cells in $\mathcal{A}$ the affinity a increases by a factor r, termed regeneration factor ($r \geq 1$), in each time step until an upper limit $a_{\max}$ is reached. Since cells in $\mathcal{A}$ are assumed to be quiescent (G0 phase), the cell cycle clock c is not updated. In contrast, cells in Ω decrease their affinity parameter a by a factor $1/d$ in which d is the differentiation constant ($d > 1$). The cell cycle clock c is increased by one. If the cell cycle clock c equals the length of the cell cycle τ_c, a division is performed, in which the parental cell is replaced by two identical copies of it. Only the cell cycle clock has been reset to $c = 0$.

The phases of the cell cycle are defined in terms of the cell cycle clock c. For cells in Ω a G1 phase of time τ_{G1} is followed by a S phase of time τ_S and a G2/M phase of time $\tau_{G2/M}$. Generally it holds $\tau_c = \tau_{G1} + \tau_S + \tau_{G2/M}$. Transition from Ω to $\mathcal{A}$ is restricted to cells in G1 phase. However, changing back into Ω the cell cycle clock c is set to the beginning of S phase.

The differentiating cells with $a < a_{\min}$ remain in Ω, and continue to divide throughout a proliferative phase with a characteristic cell cycle time, followed by a precluding maturation phase without further amplification. Finally, mature cells are removed from the system to reflect their limited life span.

Further details about the numerical implementation are provided in [47] as well as in the Appendix B.

The Roeder and Loeffler model is a successful example of how an flexible and reactive view on cellular development can be put into a mathematical context and provide insight into potential regulations on the cellular as well as on the tissue level. The model has been successfully applied to phenomena of clonal competition [221], asymmetric stem cell fates [183] and chronic myeloid leukemia [220].

From a conceptual perspective there are two limitations of the model. First, the transition functions for each cell are based on a global information (the absolute cell number within each signaling context) rather than on a local density information. However, it could be shown by d'Inverno and Saunders (publication in revision) that within a spatial structure the global information can be exchanged by a local one (based on the local concentration of a niche-derived marker substance) without any general impact on the system dynamics. Second, although the model accounts for the output of differentiating cells the mechanisms of lineage specification remain illusive. Within the next chapter a mathematical model for the description of lineage specification is presented which also puts the processes of maintenance of repopulation ability and multipotency into a common context.

4. Methods I: Modeling the lineage specification of HSCs

Within this chapter a catalog of experimentally and conceptual criteria is formulated which is subsequently used to motivate a new model of lineage specification. Generally the model should

- describe lineage specification as a dynamically regulated, temporally extended and potentially reversible process that integrates in the concept of a self-organizing stem cell system.
- extend the existing model of hematopoietic stem cell self-renewal and differentiation to incorporate the aspects of lineage specification.
- consistently explain a variety of experimentally observed phenomena.
- be as simple as possible.

After specification of the model assumptions, a formal description of the mathematical representation is provided. A sensitivity analysis of the model and a comparison to experimental results is provided in Chapter 5.

4.1. Catalog of critical phenomena

Based on the phenomenology of lineage specification introduced in Chapter 1 as well as the biological and conceptual background outlined in Chapters 2 and 3, the central aspects of the biological process can be summarized into the following catalog of criteria that need to be covered by the proposed model:

Lineage commitment (C1).

- Starting from multipotent stem or progenitor cell, lineage specification is characterized as a progressive restriction in lineage potential, finally giving rise to functionally restricted, mature cells (cf. Section 2.3.1).
- Experimental results about the lineage contribution of single hematopoietic progenitor cells [26, 222] serve as a quantitative measure for this criterion.

Generation of diversity (C2).

- Different mature cell types are generated through the process of lineage specification. Candidate mechanisms for the generation of this phenotypic diversity are cell intrinsic fluctuations caused by different sources of noise (cf. Sections 2.3.1 and 3.3.3).
- Phenotypic diversity is estimated using experimental results about the lineage contribution of paired hematopoietic progenitor cells [27, 222].

Temporal extension and reversibility (C3).

- The process of lineage specification is not an "all-or-none" decision but is characterized as a progressive restriction of lineage potential with temporal extension. This includes aspects of reversibility which, depending on the state of commitment and the particular environmental conditions, may be more or less likely (cf. Sections 2.3.2 and 2.3.3).
- Time course data on the *in vitro* differentiation of a particular mouse cell line is used to study these temporal effects.

Regulation of lineage specification (C4).

- Lineage specification is regulated on the level of individual commitment decisions (*instructive*) or by means of selective survival of preferred lineages (*selective*). Such an adaptive regulation is required on the population level for both scenarios even though individual lineage decisions might not be predictable on the single cell level (cf. Section 2.3.2).
- The aspects are discussed using data on the *in vitro* differentiation of a mouse cell line.

Priming (C5).

- On the molecular level, the uncommitted state is characterized by a coexpression of many, potentially antagonistic genes and transcription factors (cf. Section 2.3.2). In contrast, differentiated cell types show a certain typical gene expression pattern supporting their functional requirements. In this sense, lineage specification is perceived as a process that shifts gene expression from the undifferentiated coexpression state towards the dominance of a certain pattern.
- Due to the phenotypic perspective of the model these aspects are only verified in a qualitative manner.

4.2. The intracellular lineage specification model

The maintenance of HSCs is largely governed by the action of hematopoietic niches. It has been illustrated in Section 2.2 that the protective action of a particular spatial micro-environment preserves a fraction of cells in an uncommitted, mostly quiescent state with exceptionally high repopulation potential. In contrast, cells that are not under the tight regulation of the hematopoietic niches progressively lose their repopulation potential and undergo terminal differentiation.

On a conceptual level it has been shown by Roeder and Loeffler that the interplay of two, rather antagonistic environments is sufficient to explain the balance of stem cell maintenance and differentiation in the context of a self-organizing system (cf. Section 3.4). Whereas one such environment promotes stem cell maintenance the other one promotes differentiation and the loss of stem cell function. It is the adaptive ability of the cells to change between the two environments that finally leads to the establishment of a dynamically stabilized stem cell pool which also gives rise to a population of differentiating cells.

Coming back to the observation that stem cells are defined as uncommitted, multipotent cells with high regenerative potential (cf. Section 2.1) it is also commonly perceived that the lineage commitment of HSCs is closely correlated with the loss of regenerative potential. Although this correlation does by no way imply a functional identity between lineage commitment and loss of regenerative potential, it suggests that the two distinct processes are at least regulated by correlated mechanisms. Following this line of argument it is reasonable to assume that not only the process of stem cell self-maintenance but also the process of maintenance of multipotency can be explained in the context of two antagonistic environments. For the particular situation this means that the two micro-environmental configurations impose different control regimes on the process of lineage specification: whereas a *regressive control regime* maintains multipotency it is the role of a *dissipative control regime* to facilitate lineage commitment.

The antagonistic action of a *regressive* and a *dissipative control regime* is illustrated in Figure 4.1, in which the complex interaction of potentially relevant genes and gene products are depicted by a simple, intracellular network graph. Assuming that the uncommitted state is characterized by the low-level coexpression of many, potentially antagonistic genes, lineage commitment is characterized by the up-regulation of a certain lineage specific subset of genes while others are down-regulated. In this sense, the coexpression is maintained in the *regressive control regime*, whereas the dominance of one subset of genes is established in the *dissipative control regime*. Changes between the control regimes are required in order to allow for an adaptive and dynamically stabilized system. However, such changes do also imply that reversible developments in the process of lineage specification are generally possible.

A number of simplifications are necessary to reduce the complexity of the un-

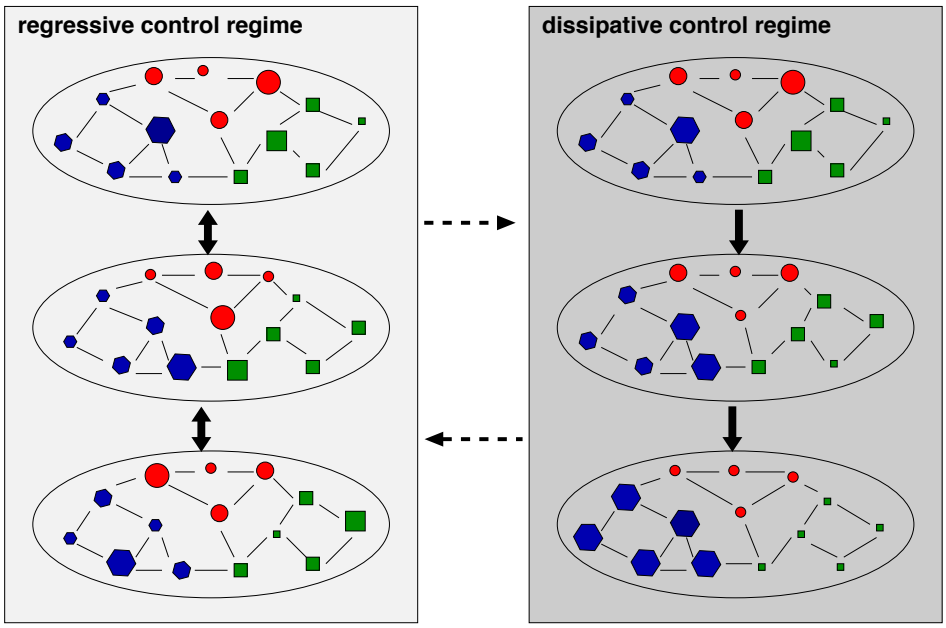

Figure 4.1.: Network dynamics under two control regimes.
Simplified representations of the state space S are depicted for a particular cell (illustrated by an intracellular interaction network). Different, coregulated components (such as genes, transcription factors, receptors) that are specific for one lineage are depicted with the same color (symbol size corresponds to the relative expression). In the *dissipative control regime* there is a clear tendency for the upregulation of one (in this case the "blue") coregulated cluster at the expense of the others (directed movement in the state space). In contrast, the *regressive control regime* supports a leveling of the expression states of the three clusters with moderate fluctuations and reversible changes (undirected movement in the state space). The cells obey the principle ability to change between the two control regimes (dashed arrows).

derlying interaction network governing the complex molecular dynamics of lineage specification and translating it into a mathematically feasible problem. Primarily, it is assumed that all signals and regulators that are specific for one common lineage fate are integrated in a single generic measure, called *lineage propensity*. This means that the action of different control regimes is not described on the level of individual genes but summarized as the combined expression of coregulated, lineage specific regulatory components (i.e. genes, transcription factors, surface factors, receptors and signaling pathways) that are representative for a certain cell fate. The level of the particular lineage propensity represents the potential of a cell to develop into the corresponding lineage. The scheme in Figure 4.1 translates into a simpler version as depicted in Figure 4.2.

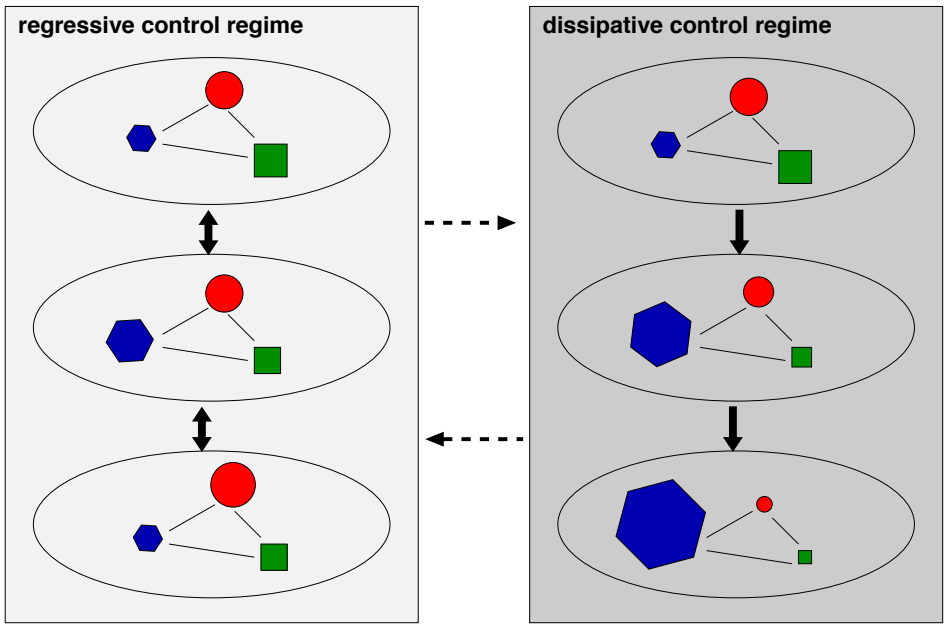

Figure 4.2.: Simplified network dynamics.
Coregulated clusters of molecular components are summarized in a single, generic measure termed *lineage propensity* (indicated by the symbols). The overall dynamics within this low-dimensional state space representation remain the same as in Figure 4.1.

It is now possible to visualize the action of the control regimes in a simplified state space diagram (Figure 4.3). For the example of three competing lineages, which are assigned to the axis of a 3-dimensional diagram, each point in the state space refers to a triplet of lineage propensities. The temporal development is consequently depicted by a trajectory as a sequence of such points. In this state space, the *regressive control regimes* leads to trajectories close to the origin, characteristic for low and balanced lineage propensities, whereas the *dissipative control regime* drives the trajectory along one of the axis, indicating a dominating lineage propensity while the others are suppressed.

Although extensive effort has been made to reveal the interaction between certain competing lineage specific genes and gene products many connections remain vaguely characterized and the resulting network dynamic can only be perceived on a very coarse scale. Therefore it is proposed to apply a simple interaction process acting on the level of the lineage propensities in order to understand the phenotypic changes in the course of cellular differentiation. In particular, it is assumed that the interaction is governed by a mutual competition process between all such propensities in which every gain (or loss) in a particular lineage propensity will

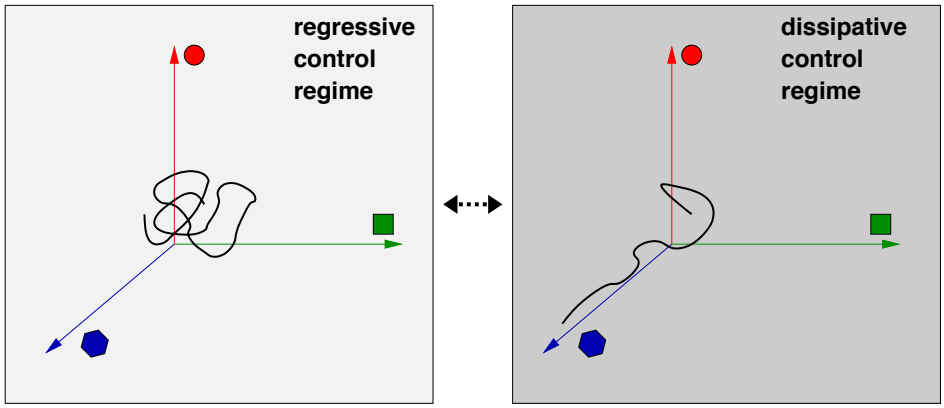

Figure 4.3.: Simplified network dynamics in a state space diagram.
In the state space diagram each dimension corresponds to one lineage propensity. A possible trajectory is indicated by the black line. Whereas in the *regressive control regime* the propensities are kept at an almost equal level, the *dissipative control regime* supports the dominance of one of them. This is an alternative representation to Figures 4.1 and 4.2.

lead to the reduction (or increase) of the remaining propensities. In order to account for the complexity of lineage development the competition process involves a stochastic element. This implies that the precise outcome of a particular lineage specification process is unpredictable. However, the overall fractions of cells committed to the cell fates in questions can still be manipulated on the level of cell populations.

The mutual competition process can act in two distinct facets to represent the different actions associated with the control regimes outlined above. In simple words, it is assumed that in the regressive control regime the competition process leads to the maintenance of the low-level coexpression (priming) by penalizing deviations from the mean expression level, while in the dissipative control regime the dominance of one or the other lineage propensity is promoted (commitment) by enhancing the occurring deviations.

The above arguments lead to a catalog of formal assumptions on which the mathematical formalization is based:

- Coregulated, lineage specific components (e.g. genes, transcription factors, receptors) are summarized into *lineage propensities*. Each lineage propensity represents the potential of a cell to develop into the corresponding lineage.
- Interactions between the lineage propensities are represented by a stochastic competition process with different configurations for the *regressive* and the *dissipative control regime*.

- The lineage propensities are described as relative values (normalized to unity). This means that the gain (or loss) in a particular lineage propensity will lead to the reduction (or increase) of the remaining propensities.

- The lineage propensities represent cellular features that are inherited to the daughter cells. Inheritance of the lineage propensities is symmetric to both the daughter cells.

- The outcome of a particular lineage specification process can only be predicted in a probabilistic sense. The overall outcome on the cell population level can be either regulated by modifying parameters of the intrinsic competition process or by a positive selection for cells with certain patterns of their lineage propensities.

4.3. Mathematical representation of the intracellular model

A mathematical formalization is necessary in order to study the quantitative model behavior under the set of stated assumptions.

4.3.1. Lineage specification as a cellular property

The lineage specification state of a model cell at any given time point is characterized by the actual levels of a number N of different lineage propensities denoted by a vector $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_N(t))$. N represents the number of different lineages in which the cell can potentially differentiate. The propensities x_i, which can take values between zero and one, represent relative propensity levels for the development into the N possible lineages. In other words, the lineage propensity vector $\mathbf{x}(t)$ is always normalized to 1 ($\sum_i x_i(t) = 1$).

4.3.2. Dynamics of the competition process

The stochastic competition process, which is modeled on a discrete time scale (typically measuring one hour), is organized as a three step process.

- In the **first step of the competition process**, one lineage propensity x_i is randomly chosen for the update procedure. Herein, the probability P for choosing a particular lineage i equals its propensity $x_i(t)$:

$$P(i) = x_i(t) \tag{4.1}$$

- In the subsequent, **second step of the competition process** the chosen lineage propensity $x_i(t)$ is updated according to

$$x_i(t+1) = x_i(t)(1+m_i) \tag{4.2}$$

whereas the other propensities remain unchanged in the first step. The general idea of this update procedure is inspired by a so called *Pólya* urn model (see Appendix A.1).

The lineage specific reward for the randomly chosen lineage i is defined as a function $m_i = f_m(x_i(t))$. For all other lineages j the reward is set to $m_j = 0$. The parameterization of the particular reward functions for each individual lineage is the primary target to manipulate the overall system dynamics. For the update, two different control regimes are assumed depending on the micro-environmental context the cell is actually exposed to and which are characterized by qualitatively different reward functions:

Regressive control regime: Deviations between the chosen lineage propensities x_i and a common mean propensity level are modulated by positive or negative rewards m_i, respectively. The reward function can either be described by a linear function of the type

$$m_i = f_m(x_i(t)) = b_i x_i + n_i \tag{4.3}$$

with negative slope ($b_i < 0$) and a root at $x^R = -n_i/b_i$ or, alternatively, by a sigmoid function of type

$$m_i = f_m(x_i(t)) = \frac{-\sigma_i s_i (x_i - x^R)}{\sqrt{1 + (s_i(x_i - x^R))^2}}\,. \tag{4.4}$$

This sigmoid function is defined by the root x^R, the steepness parameter s_i and the saturation parameter σ_i. The root at x^R should in both cases be chosen such that $x^R \approx \bar{x} = 1/N$ in order to get convergence to the mean propensity level $\bar{x} = 1/N$. In this setting the reward m_i is positive for $x_i < x^R$ and negative for $x_i > x^R$. Examples of the reward functions are shown in Figure 4.4(A).

The linear reward function in equation (4.3) is suitable in many cases. However, for large deviations of the chosen lineage propensity x_i from the mean propensity level $\bar{x} = 1/N$ the reward could potentially fall below $m_i = -1$. This, in turn, leads to negative lineage propensities in the update function in equation (4.2) which are not defined. In order to prevent this case a lower limit for the linear reward has been introduced at $m_i = -0.5$.

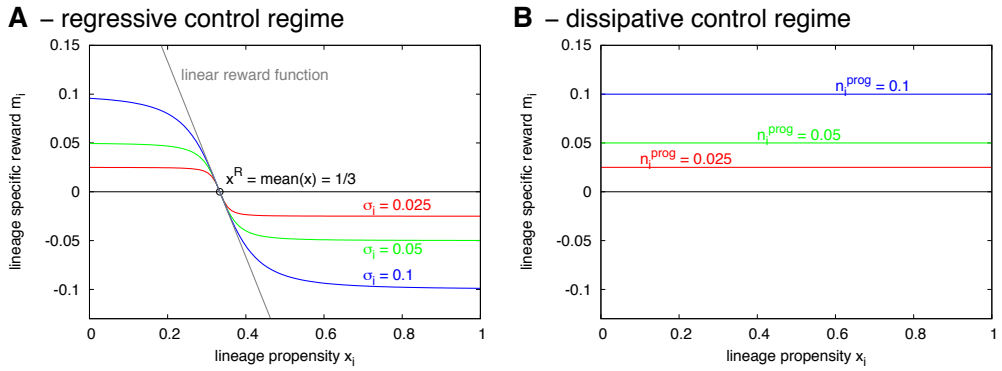

Figure 4.4.: Reward functions $m_i = f_m(x_i(t))$ for the two control regimes. **(A).** For the regressive control regime a linear reward function (given in equation (4.3)) with slope $b_i = -1$ and three examples of sigmoid reward functions (given in equation (4.4)) are shown with parameters $x^R = 1/3$, $s_i = 1/\sigma_i$ and $\sigma_i = 0.1, 0.05, 0.025$. **(B).** Three examples of linear reward functions (equation (4.5)) are shown for the dissipative control regime. Parameters are $n_i = 0.1, 0.05, 0.025$.

Alternatively, an appropriate sigmoid function can be used as given in (4.4). Particular examples are shown in Figure 4.4(A). Taking the second derivative of equation (4.4), it can be shown that the maximum slope (occurring at the root point x^R) is given as $b_i = -\sigma_i s_i$. For the three examples illustrated in Figures 4.4 the steepness is set to $s_i = 1/\sigma_i$. It follows that the slope for all three examples is $b_i = -1$ at the root point x^R. Therefore the functions differ only in their saturation level σ_i.

Dissipative control regime: The chosen lineage propensity x_i is always enhanced by lineage specific, positive rewards m_i. In the simplest case a constant reward is chosen, resulting in a linear reward functions of the type

$$m_i = f_m(x_i(t)) = n_i \tag{4.5}$$

These functions have zero slope and are restricted to the positive plane ($n_i > 0$). This way $m_i \equiv n_i$ is independent of x_i (Figure 4.4(B)). This leads to increasing divergence from the mean propensity level $\bar{x} = 1/N$. The preferential update of lineages with high propensities (due to the coupling of update probability and actual lineage propensity in the first update step), ultimately leads to the dominance of one lineage over the others.

- In the final, **third step of the competition process**, the lineage propensity vector $\mathbf{x}(t)$ is normalized to unity. This normalization accounts for the

concept of antagonistic interaction between different lineages: if one propensity is up-regulated, the other ones are (relatively) down-regulated and vice versa. It can be shown mathematically that this renormalization can be interpreted as a general decay term acting on all lineage propensities. For further details the reader is referred to Appendix A.2.

Taken together, the complete update procedure reads as:

$$\mathbf{x}(t+1) = \frac{1}{C^n}\mathbf{x}(t)(1+\bar{m}(t)) \tag{4.6}$$

in which $\bar{m}(t)$ corresponds to the vector of lineage specific rewards which is recalculated at every time step t. The individual values are generally set to $m_j = 0$ except for the updated lineage i which has been chosen with probability $P(i) = x_i(t)$. For the updated lineage $m_i = f_m(x_i(t))$ is calculated according to the corresponding control regime as described in equations (4.3) to (4.5). The normalization constant C^n accounts for the subsequent normalization of the lineage propensity vector $\mathbf{x}$ to unity. In particular, C^n is given as

$$C^n = \sum_{j\in\{1\ldots N\}} x_j(t)(1+m_j) = 1 + m_i \tag{4.7}$$

in which m_i corresponds to the individual reward of the updated lineage i.

Correlations between lineages. As outlined in Section 2.3.1 certain lineage fates are more closely related than others. For example, it has been observed that the development of granulocytes and macrophages is often closely correlated, whereas similar effects are less frequent among other cell types. Since this phenomenon is caused by common developmental pathways, such a correlation in the modeling approach is reflected by a functional coupling of the particular lineage propensities. Technically this is implemented as follows: Given there is a correlation between any two lineages propensities, say i and j, the update procedure is extended. If lineage i is chosen for the update process, x_i is updated according to the update function in equation (4.2). However, at the same instance also the correlated lineage j undergoes a modified update procedure: the propensity x_j is simultaneously updated according to $x_j(t+1) = x_j(t)(1+\gamma_{ij}m_j)$. The parameters γ_{ij} characterize the correlation and are considered between $-1 < \gamma_{ij} < 1$ (parameter values $\gamma_{ij} < 0$ account for negative correlations). A particular application is provided in Section 5.4.1.

4.3.3. Phenotypic mapping

In the model one needs to decide how changes of the intracellular state correspond to changes in the phenotype of the cells. Such a mapping is necessary to compare

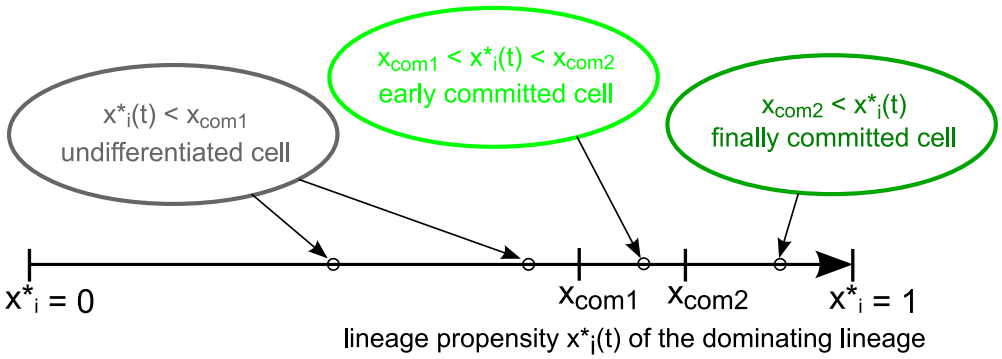

Figure 4.5.: Phenotypic mapping.
The bar indicates the lineage propensity x_i^* of the *dominating lineage*. According to this value and the threshold values $x_{\mathrm{com1/2}}$ the cell is classified as uncommitted, early or finally committed cell.

the results of the modeling approach to experimental data which typically detect phenotypic changes during the course of differentiation.

For the particular mapping, it is assumed that the phenotype of an individual cell is characterized by its highest actual lineage propensity $x_{\max} = \max_i x_i$. The lineage i for which holds $x_i(t) = x_{\max}$ is termed the *dominant lineage*. Subsequently this propensity is denoted as $x_i^*(t)$. For a dominant lineage with $x_i^*(t) < 0.5$ the assignment should be handled with care since potentially more than one lineage propensity can have closely similar values.

According to the lineage propensity of the dominant lineage $x_i^*(t)$ the cells are classified in comparison to a threshold value x_{com1} as undifferentiated cells ($x_i^*(t) < x_{\mathrm{com1}}$) or as cells committed to lineage i ($x_i^*(t) > x_{\mathrm{com1}}$). If further states of commitment can be defined phenotypically it might be useful to introduce a further class of cells with a second threshold value x_{com2}. The simple sketch in Figure 4.5 illustrates such an example with a secondary distinction in early and late committed cells.

It should be clearly stated that the specification of the threshold values $x_{\mathrm{com1/2}}$ is used solely for the phenotypic mapping of model results to experimental data and does not imply the irreversibility of the commitment decision. However, the probability for a change of the dominant lineage decreases with increasing values of the propensity x_i (see analysis in Section 5.1.2).

4.3.4. Regulation of lineage specification

Conceptually there are two different mechanisms for the regulation of the outcome of a lineage specification process. Whereas the biological background has been discussed in Section 2.3.2, a conceptual placement is provided below.

Lineage specification as an instructive process. The shape of the reward functions f_m defined in equations (4.3) to (4.5) is a suitable parameter for the regulation of the lineage specification process. In particular, the constant reward in the dissipative regime $f_m(x_i(t)) = m_i = n_i$ is especially sensitive to influence the final lineage commitment as the choice of different values n_i for different lineages i skews the probability for their occurrence. A detailed sensitivity analysis is provided in Section 5.1 of the next chapter.

Although, cell death might occur in this scenario, it generally does not have a regulating function and affects all cells in a similar fashion (i.e. *background cell death* with equal intensity). Therefore, the regulation of lineage specification is solely attributed to the manipulation of the reward functions f_m. As this is considered to be an instructive process, mediated e.g. by the action of lineage specific cytokines, the whole process is referred to as *instructive lineage specification*.

Lineage specification as a selective process. In the case that the lineage specification process outlined in Section 4.3.2 has no intrinsic skewing (i.e. the reward functions f_m are identical for all lineages i) all N possible lineages are generated with the equal probabilities. Therefore a regulation of the fraction of cells in each lineage can only be achieved by the process of selective cell death. Depending on the action of a supportive or permissive environment (e.g. lineage specific cytokines, feedback regulations) the survival of certain lineages is propagated while other lineages are not supported and undergo cell death.

As the regulation of the fraction of cells in each individual lineage is solely regulated by selective survival signals and targeted cell death, the process is referred to as *selective lineage specification*.

This conceptual approach assumes cell death occurrence in two principle meanings, namely as a general effect occurring in all cells (*background cell death*) or as an effect that is coupled to certain cellular properties (*selective cell death*). The translation into the model concept of lineage specification is outlined below:

Background cell death. Background cell death is defined as unspecific cell death process acting on all cell types irrespective of their current state.

Technically, the occurrence of cell death is modeled as a stochastic process. The probability for a cell to undergo cell death in any given update step, referred to as Φ^B, is given as:

$$\Phi^B = \phi \tag{4.8}$$

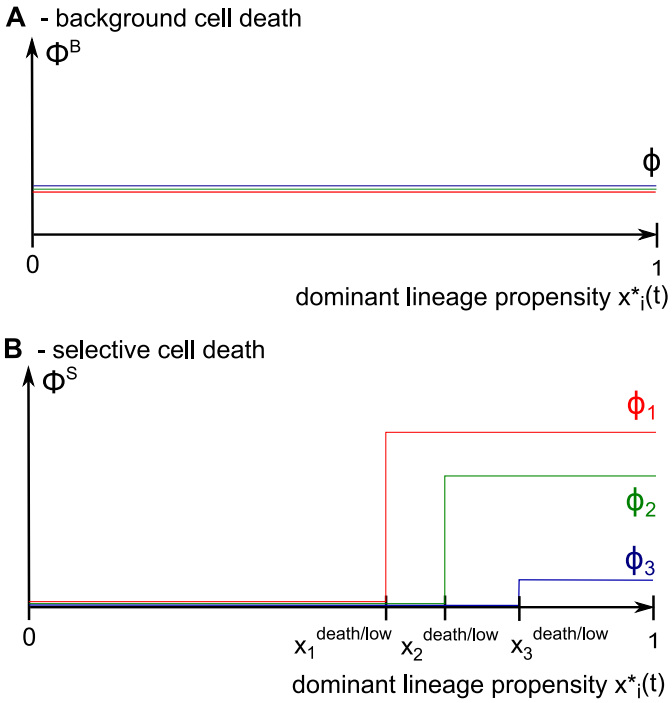

Figure 4.6.: Cell death intensities Φ.
The probability for cell death during one update step Φ is shown as a function of the dominating lineage propensity $x_i^*(t)$. **(A).** For the *background cell death* Φ^B remains constant. **(B).** For the *selective cell death* Φ^S is defined by lineage specific step functions (indicated by the colors). Without loss of generality $x_i^{\text{death/high}} = 1$ for all i.

Here, ϕ is a general and constant cell death intensity which does not depend on the lineage propensities $\mathbf{x}(t)$ (see also Figure 4.6(A)).

Selective cell death. In contrast, the selective cell death does not act equally on all cells but cells with certain dominant lineages are primary targets whereas cells with other dominant lineages are spared. In this sense, the occurrence of cell death for a particular cell depends explicitly on its actual lineage propensities $\mathbf{x}(t)$.

Again, the occurrence of cell death is modeled as a stochastic process. However, the probability $\Phi^S(t)$ that a cell undergoes cell death in a given time step t is now described as function of the dominating lineage propensity $x_i^*(t)$. In the most simple approach $\Phi^S(t)$ is defined as a step function of the form

$$\Phi^S(t) = \begin{cases} \phi_i & \text{for } (x_i^{\text{death/low}} < x_i^*(t) < x_i^{\text{death/high}}) \\ 0 & \text{else} \end{cases} \tag{4.9}$$

In this case, the probabilities ϕ_i as well as the boundaries $x_i^{\text{death/low}}$ and $x_i^{\text{death/high}}$ are individual parameters for each lineage i. Choosing a higher value of $\phi_i > \phi_j$ leads to an increased cell death probability for cells with the dominating lineage i as compared to cells with dominating lineage j. In this sense, it is the propensity of the dominating lineage x_i^* that further specifies equation (4.9). Typical curves for the cell death intensities Φ^S are provided in Figure 4.6(B).

Practically, the lower bound for the action of the selective cell death is chosen such that $x_i^{\text{death/low}} > 0.5$. Only in this case the definition of a dominant lineage is robust against small perturbations [1]. Similar restrictions apply for the upper bound $x_i^{\text{death/high}}$. As it is unlikely that an initially unsupported lineage has better survival chances later in development, the upper bound for the action of the selective cell death is generally set to $x_i^{\text{death/high}} = 1$.

Integrating the concepts of cell death in the mathematical and numerical formulation of the lineage specification model, the three-part update procedure introduced in Section 4.3.2 is complemented by one of the above scenarios for the occurrence of cell death. Consequently the scenarios of *instructive* and *selective lineage specification* are summarized as follows:

Instructive lineage specification.

- lineage specific reward functions $f_m(x_i)$ skew the intrinsic lineage decision

  and

- background cell death with intensity Φ^B acts on all cells without regulation of the lineage contribution

[1] For any lineage propensity $x_i < 0.5$ two or more lineages can have closely similar values, even fluctuating around a certain common mean value. In this case the dominant lineage would change frequently between subsequent update steps.

Selective lineage specification.

- identical reward functions $f_m(x_i)$ for all cells lead to balanced lineage decisions

  and

- selective cell death with intensity Φ^S regulates the final lineage contribution

4.4. Integration of lineage specification in the model of HSC self-maintenance

The intracellular lineage specification dynamics described above are defined under the control of two antagonistic control regimes. Following the stated criteria that reversible developments are generally possible, transitions between the regressive and the dissipative control regime (and vice versa) need to be included in a comprehensive model of HSC organization.

The previously proposed model of HSC organization introduced by Roeder and Loeffler (compare Section 3.4) describes adaptive and self-organizing HSC self-renewal in the context two, structurally different and antagonistic micro-environmental conditions: one promoting cellular quiescence and maintenance of the self-renewal ability (signaling context $\mathcal{A}$), the other promoting proliferation and loss of self-renewal ability (signaling context Ω).

It appears as a logical consequence to extend this idea of micro-environmentally directed development on the dynamics of lineage specification. In particular, it is proposed that the two signaling contexts $\mathcal{A}$ and Ω also impose contrary effects on the intracellular lineage specification dynamics of each individual cell by

- Coupling the *dissipative control regime* to *signaling context* Ω.

  and

- Coupling the *regressive control regime* to *signaling context* $\mathcal{A}$.

Induced by this superposition of regulating mechanisms, cells in $\mathcal{A}$ preserve the uncommitted state and simultaneously maintain their ability to act as stem cells. Furthermore, the processes of lineage commitment is correlated with the loss of self-renewal ability in signaling context Ω.

Technically, the cellular properties defined in the original model of HSC self-renewal outlined in Section 3.4 (i.e. attachment affinity a, position in the cell cycle

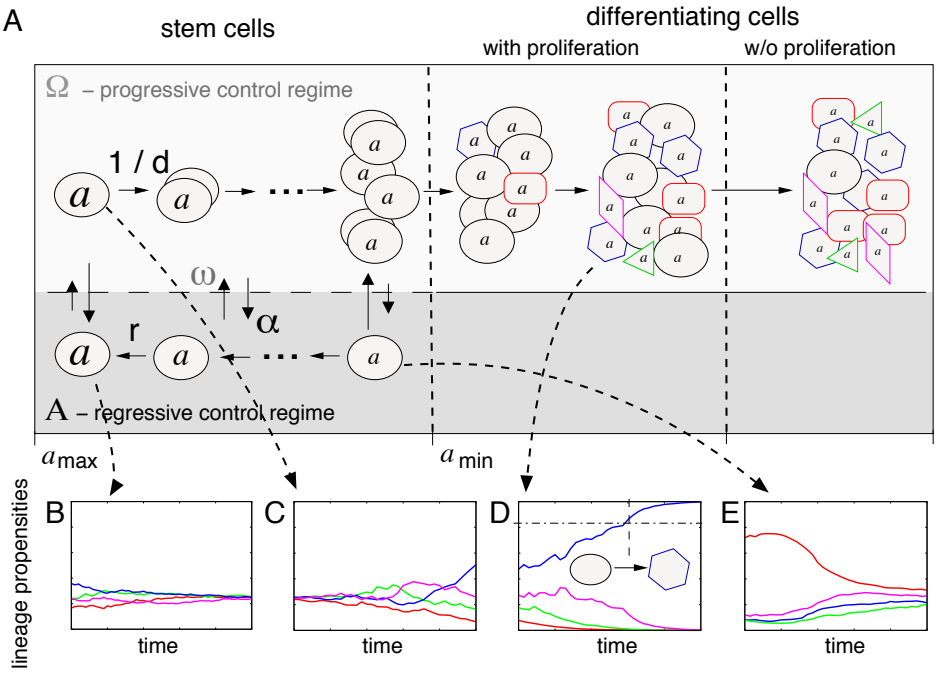

Figure 4.7.: Extended model concept.
(A). The general model layout and the parameters correspond to Figure 3.12. Upon integration of the lineage specification dynamics the cells now acquire a particular lineage during differentiation in Ω (indicated by the change in shape and color). **(B-E).** Time courses of the lineage propensities $\mathbf{x}$ for individual cells of interest. **(B).** The non-proliferative, stem cell supporting signaling context $\mathcal{A}$ hosts multipotent cells which are characterized by the balanced low level co-expression of the lineage propensities. **(C).** Due to dissipative control regime in signaling context Ω the balanced coexpression is upset and one lineage propensity is expanded at the cost of the others. **(D).** Continuation of the process in which the particular factor manifests the lineage decision and identifies the cell as committed ($x_i^* > x_{\text{com}}$). **(E).** Intracellular development of a cell which has been recaptured into signaling context $\mathcal{A}$. Here, the regressive control regime counteracts the differentiation process and reestablishes the typical priming pattern.

c and the affiliation to either the signaling context $\mathcal{A}$ or Ω) are now extended to incorporate the vector of lineage propensities $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_N(t))$. During the sequential processing of each cell at the discrete time steps of the original model, the cell's lineage propensities $\mathbf{x}$ are additionally updated according to the dynamics of lineage specification outlined in Section 4.3. Depending on the current signaling context $\mathcal{A}$ or Ω, either the regressive or the dissipative control regime is applied for the update procedure.

The transitions between signaling contexts $\mathcal{A}$ and Ω (and vice versa) are still

governed by the functions defined in equations (3.1) - (3.3). These functions depend only on the number of cells in the target signaling context and on the cell's individual affinity a, but they do not explicitly depend on the cell's lineage propensity $\mathbf{x}(t)$. However, as the dynamics of lineage specification are functionally integrated in the model of HSC self-renewal as outlined above, there is a close correlation between a cell's affinity a and its lineage propensity $\mathbf{x}$.

A comprehensive illustration of the extended model concept is shown in Figure 4.7. As the intrinsic dynamics of lineage specification are difficult to illustrate, typical time courses of the lineage propensities $\mathbf{x}$ for individual cells of interest are shown below the main graphic.

Further details of the necessary changes to the numerical program structure of the Roeder and Loeffler model are provided in Appendix B. For details on the implementation of the novel model of lineage specification the reader is referred to Appendix C.

5. Results I: Dynamics of lineage specification

The first part of this chapter provides a general overview of the dynamic behavior of the proposed model of lineage specification including a detailed description of the two control regimes and the system's response to parameter changes. The second part presents a range of experimental situations to which the model has been applied.

5.1. Lineage specification in the regressive and the dissipative control regime

In order to understand the rich dynamical features of the lineage specification model it is necessary to illustrate the behavior for the regressive and for the dissipative control regime separately. An integrated view of the extended stem cell model including both the aspects of self-renewal and lineage specification is provided at the end of the section.

5.1.1. Dynamics in the regressive control regime

General behavior In the regressive control regime the system is characterized by a convergence of the lineage propensities x_i towards their mean propensity level $\bar{x} = 1/N$. This is facilitated by the reward functions (given in equations (4.3) and (4.4)) which penalize deviations from the mean propensity level. Given that the root of the reward functions x^R coincides with the mean propensity level $\bar{x} = 1/N$, a positive reward ($m_i > 0$) is imposed in case that the propensity of the chosen lineage is below the mean level of the root $x_i < x^R = \bar{x}$ and a negative reward ($m_i < 0$) is imposed in the contrary case ($x_i > x^R = \bar{x}$). Following the update procedure outlined in Section 4.3.2 the individual propensities approach the mean propensity level $x_i \to \bar{x} = 1/N$ with increasing time. Typical examples of the resulting trajectories are shown in Figure 5.1 in which the dynamics of convergence are shown for different independent realizations.

Starting from a system in which one lineage propensity dominates over the others, it appears that the convergence towards the mean coexpression level $\bar{x}$ is a suitable measure to characterize the dynamics of the lineage propensities $\mathbf{x}$ in the regressive control regime. Technically, for each individual trajectory the time point is recorded when a lineage propensity x_i passes a certain threshold x^t. Time averages for different individual threshold values are obtained separately for each lineage i. A so called *convergence curve* is obtained by connecting these

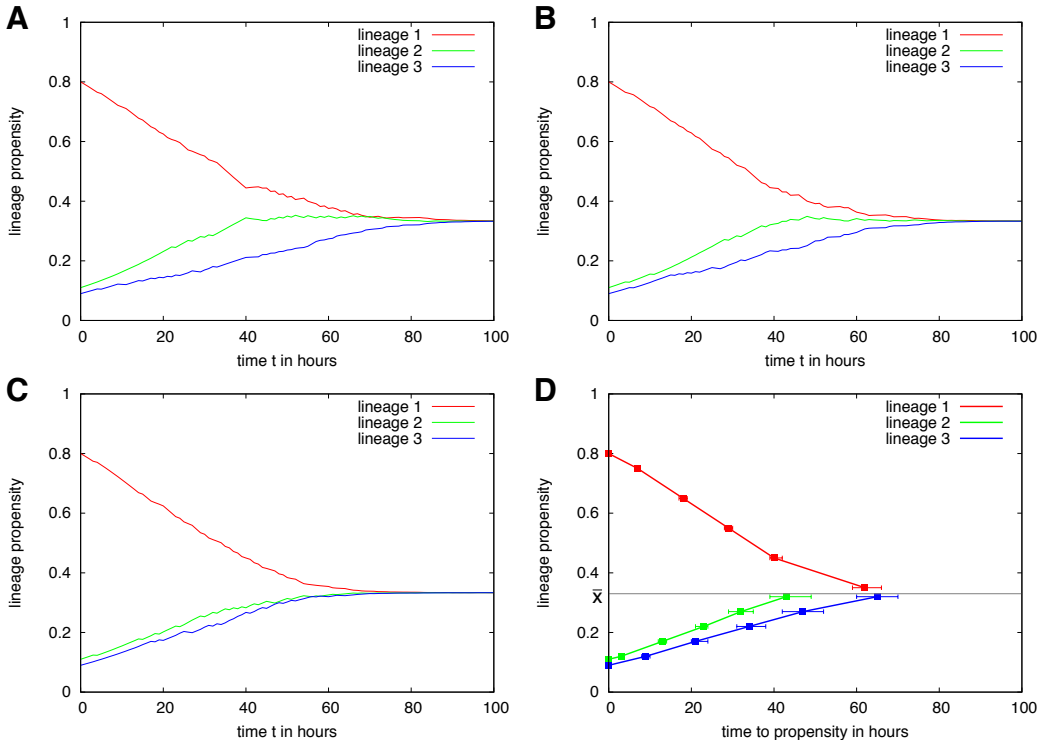

Figure 5.1.: Trajectories in the regressive control regime.
(A - C). Typical trajectories (lineage propensity x_i vs. time) for a system of three interacting lineages with initial values $\mathbf{x}(t=0) = \{0.8, 0.11, 0.09\}$. Parameters of the sigmoid reward function in equation (4.4) are set to $\sigma_i = 0.05$, $s_i = 20$, and $x_i^R = 1/3$ for $i = 1, 2, 3$. **(D).** Characteristic convergence curves averaged over 10000 independent trajectories for the dominating lineage 1 (red) and for the two competing lineages 2 (green) and 3 (blue). Threshold values for the convergence curves (indicated by the squares) are set to $x^t = \{0.75, 0.65, 0.55, 0.45, 0.35\}$ for lineage $i = 1$ and $x^t = \{0.12, 0.17, 0.22, 0.27, 0.32\}$ for lineage $i = 2, 3$. Error bars indicate the interquartile range.

time averages for several threshold values. These convergence curves characterize the averaged temporal development of the convergence process for each individual lineage i and are used throughout this section to illustrate the system behavior in the regressive control regime. Examples are shown in Figure 5.1(D) in which the characteristic convergence curves are averaged over 10000 independent simulation runs in which the initial values of the lineage propensities are set to $\mathbf{x}(t=0) = \{0.8, 0.11, 0.09\}$. Details of the simulation procedure are provided in Appendix D.1.

The speed of convergence critically depends on the shape of the reward function, but also on the number of competing lineage propensities N. In the simplified

case, that the mean propensity level $\bar{x} = 1/N$ and the root x^R of the reward function equal ($x^R = \bar{x}$), fluctuations around the mean propensity level are gradually suppressed until the system reaches a steady state $x_i = \bar{x}$, $\forall\, i$. This idealized scenario is subsequently used to outline the system behavior in the regressive control regime. However, the divergence of the root x^R of the reward function from the mean coexpression level $\bar{x} = 1/N$ leads to fluctuations of the ground state and can possibly introduce a lineage bias. The latter case is discussed in Section 5.3.

Dependence on the reward function. The shape of the reward function critically regulates the dynamics of convergence of the lineage propensities $\mathbf{x}$ towards the mean level $\bar{x}$. In particular, the class of the reward function f_m (sigmoid vs. linear), the slope b at the root point as well as the saturation level σ (in the case of a sigmoid reward function) sensitively influence the characteristic convergence curves. The dependence on each of these parameters is discussed below:

- **Classes of reward functions.** As outlined in Section 4.3.2 two different types of reward functions f_m have been studied, namely a linear (equation (4.3)) and a sigmoid one (equation (4.4)). Figures 5.2(A) and (C) show the characteristic convergence curves for different reward functions f_m shown in the corresponding Figures 5.2(B) and (D), respectively. For clear representation, only the convergence curves for the dominating lineage starting at $x_i^*(t = 0) = 0.8$ are shown.

  The sigmoid reward functions defined in equation (4.4) can be adapted such that the maximal slope b_i occurring at the root point x^R equals the slope of a corresponding linear reward function with the same root point. As shown in Figures 5.2(B) and (D) the sigmoid functions (indicated in red and green) diverge from their linear counterpart (indicated in blue) depending on their saturation parameter σ.

  Illustrated by the characteristic convergence curves in Figures 5.2(A) and (C), the linear reward functions generally lead to a faster convergence towards the mean propensity level $\bar{x}$ as compared to the sigmoid reward functions. This is caused by the fact that for propensities $x_i \gg \bar{x}$ the negative reward m_i for the linear reward function is much larger as compared to the reward for the sigmoid function. However, for lineage propensities $x_i \approx \bar{x}$ the dynamics become more similar.

  It should be noted that for large values of the saturation parameter σ the sigmoid functions and the linear functions are identical in the interval of interest $[0, 1]$. As the sigmoid reward functions appear more robust (compare Section 4.3.2) the further discussion is based on this type of function unless otherwise noted.

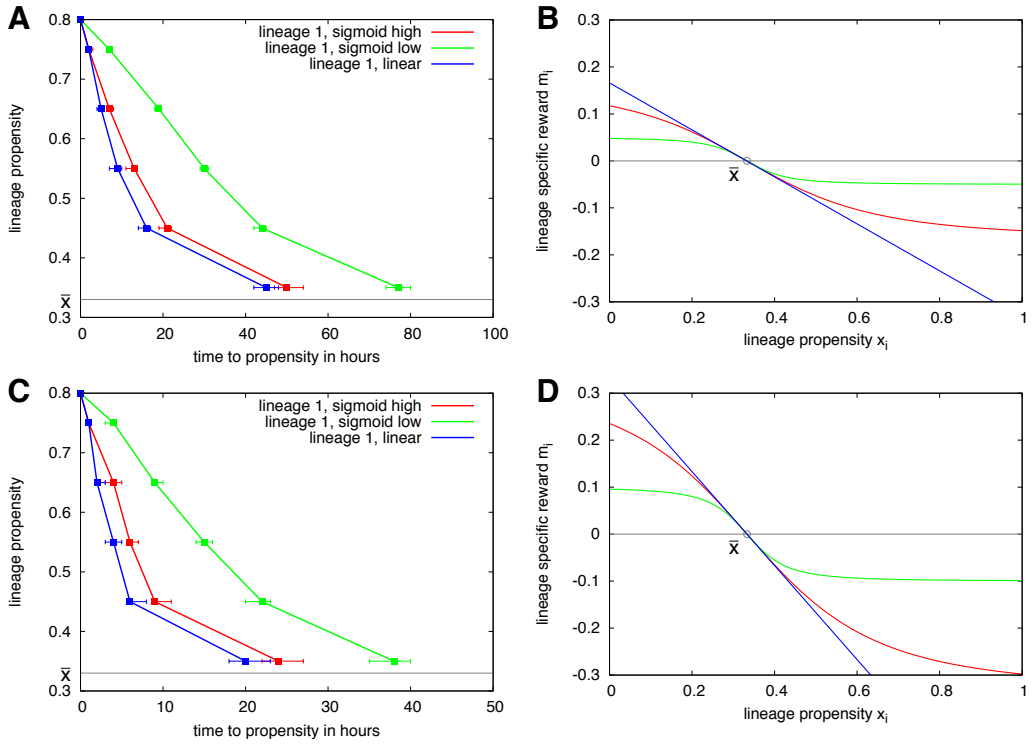

Figure 5.2.: Convergence curves for linear and sigmoid reward functions. **(A)** and **(C)** show the characteristic convergence curves (averaged over 10000 realizations initiated with $\mathbf{x}(t=0) = \{0.8, 0.11, 0.09\}$) for the dominating lineage propensity $x_i^*(t)$. The corresponding reward functions (identified by equal colors) are shown in **(B)** and **(D)**, respectively. Parameters of the sigmoid reward function shown in (B) (and thus related to the convergence curves in (A)) are set to $\sigma_i = 0.05$ (green), $\sigma_i = 0.166$ (red), $b_i = -0.5$, and $x_i^R = 1/3$ for all $i = 1, 2, 3$. Parameters of the linear reward function (blue) are $b_i = -0.5$, and $x_i^R = 1/3$ for all i. Parameters in (D) (related to (C)) are set to $\sigma_i = 0.1$ (green), $\sigma_i = 0.33$ (red), $b_i = -1$, and $x_i^R = 1/3$ for all i for the sigmoid reward function and $b_i = -1$, and $x_i^R = 1/3$ for all i for the linear reward function (blue).

- **Slope of the reward function.** The slope b_i of the sigmoid reward functions is a crucial parameter regulating how fast the system converges towards the mean propensity level $\bar{x} = 1/N$. As illustrated in Figures 5.3(A) and (B) the convergence curves are directly correlated with the steepness of the reward function as larger values of b_i lead to a faster convergence. However, as the reward functions are more similar for values $x_i \gg \bar{x}$ the difference in the convergence curves only become obvious once the lineage propensities $\mathbf{x}$ approach their mean value $\bar{x}$. In this regime ($x_i \approx \bar{x} = x^R$) the differences between the reward functions are most pronounced.

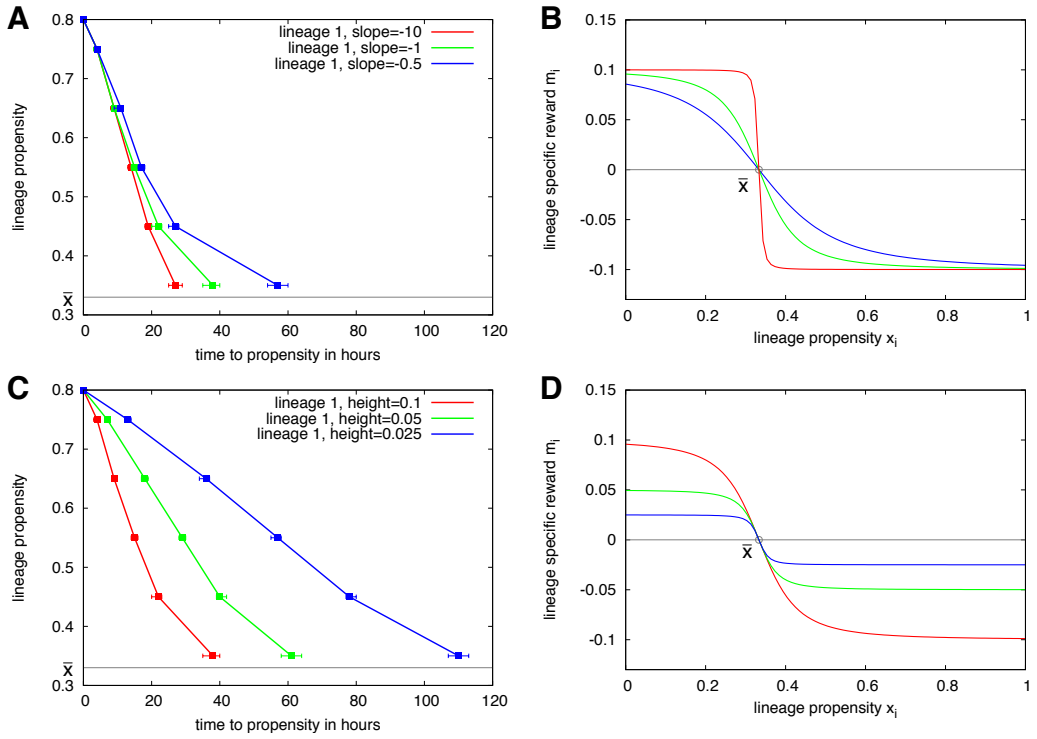

Figure 5.3.: Convergence curves depending on the slope and the saturation level of the sigmoid reward function.
(A) and **(C)** show characteristic convergence curves (averaged over 10000 realizations initiated with $\mathbf{x}(t=0) = \{0.8, 0.11, 0.09\}$) for sigmoid reward functions with different shape. The corresponding reward functions are shown in **(B)** and **(D)**, respectively. Parameters of the sigmoid reward function with different steepness in (B) (thus related to the color coded convergence curves in (A)) are set to $\sigma_i = 0.1$, $x_i^R = 1/3$, and $b_i = -0.5$ (blue), $b_i = -1$ (green), $b_i = -10$ (red) for all lineages i. For the scenario with different saturation levels σ in (D) (and related to the convergence curves in (C)), parameters are set to $\sigma_i = 0.025$ (blue), $\sigma_i = 0.05$ (green), and $\sigma_i = 0.1$ (red), and $x_i^R = 1/3$, $b_i = -1$ for all lineages i.

- **Saturation level of the reward function.** The saturation level σ of the sigmoid reward function turns out to be a similarly sensitive parameter as the slope b_i outlined above. Keeping the slope b_i constant and varying the saturation level σ_i, the graphs in Figures 5.3(C) and (D) illustrate that the convergence to the mean expression level $\bar{x}$ is accelerated for higher values of σ_i.

Within the scope of this thesis, both the slope b_i and the saturation level σ_i appear as sensitive parameters regulating the dynamical behavior of the lineage propensities $\mathbf{x}$ in the regressive control regime.

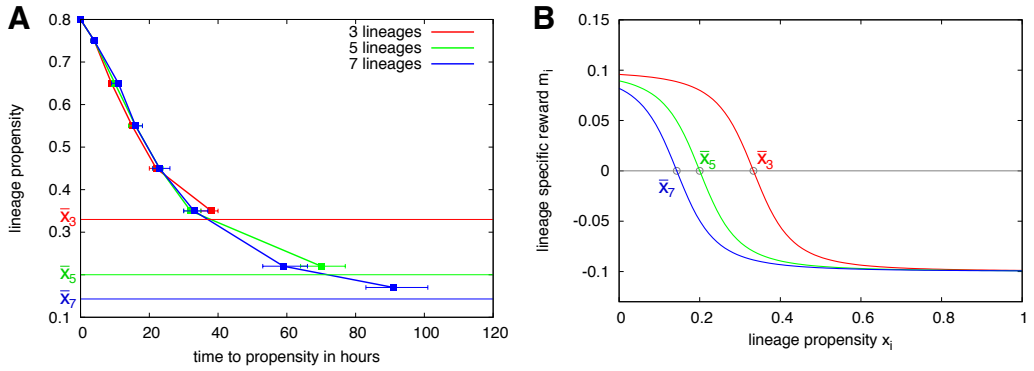

Figure 5.4.: Convergence curves depending on the number of lineages N. **(A)** shows the characteristic convergence curves for systems with different numbers of competing lineages: $N = 3$ (red) initialized with $\mathbf{x}(t = 0) = \{0.8, 0.11, 0.09\}$; $N = 5$ (green), initialized with $\mathbf{x}(t = 0) = \{0.8, 0.05, 0.05, 0.05, 0.05\}$; $N = 7$ (blue), initialized with $\mathbf{x}(t = 0) = \{0.8, 0.033, 0.033, 0.033, 0.033, 0.033, 0.033\}$. For each scenario 10000 individual realizations have been performed to calculate the shown average convergence curves. As the mean expression level $\bar{x}$ scales with $1/N$, the root x^R is adapted for the sigmoid reward functions shown in **(B)** with $x^R = 1/3$ (red), $x^R = 1/5$ (green), $x^R = 1/7$ (blue), and $\sigma_i = 0.1$, $b_i = -1$ for all lineages i.

Dependence on the number of lineages. For varying numbers of competing lineage propensities N the sigmoid reward functions need to be modified. As the mean propensity level $\bar{x} = 1/N$ scales with the number of lineages N, the root point $x^R = \bar{x}$ of the reward functions is shifted along the x-axis. This dependency is illustrated in Figure 5.4(B) showing the sigmoid reward functions for the scenarios $N = 3, 5, 7$. However, as indicated by the characteristic convergence curves in Figure 5.4(A) this shift only has a minor effect on the dynamics of the convergence towards the mean propensity level $\bar{x}$. In this sense the system appears robust against variations in the number of competing lineages.

5.1.2. Dynamics in the dissipative control regime

General behavior. In the dissipative regime, the system initially undergoes a competitive phase characterized by moderate fluctuations of the lineage propensity levels. These fluctuations occur, since all propensities are generally in the same range ($x_i \approx \bar{x} = 1/N$) and are therefore equally likely to be chosen for the update procedure. However, in the course of time one lineage propensity takes over, reducing the probability for the update procedure to act on the competing lineages (see also Section 4.3.2 and Appendix A.1). The increase of the dominating lineage propensity leads to reduced fluctuations and an acceleration of the

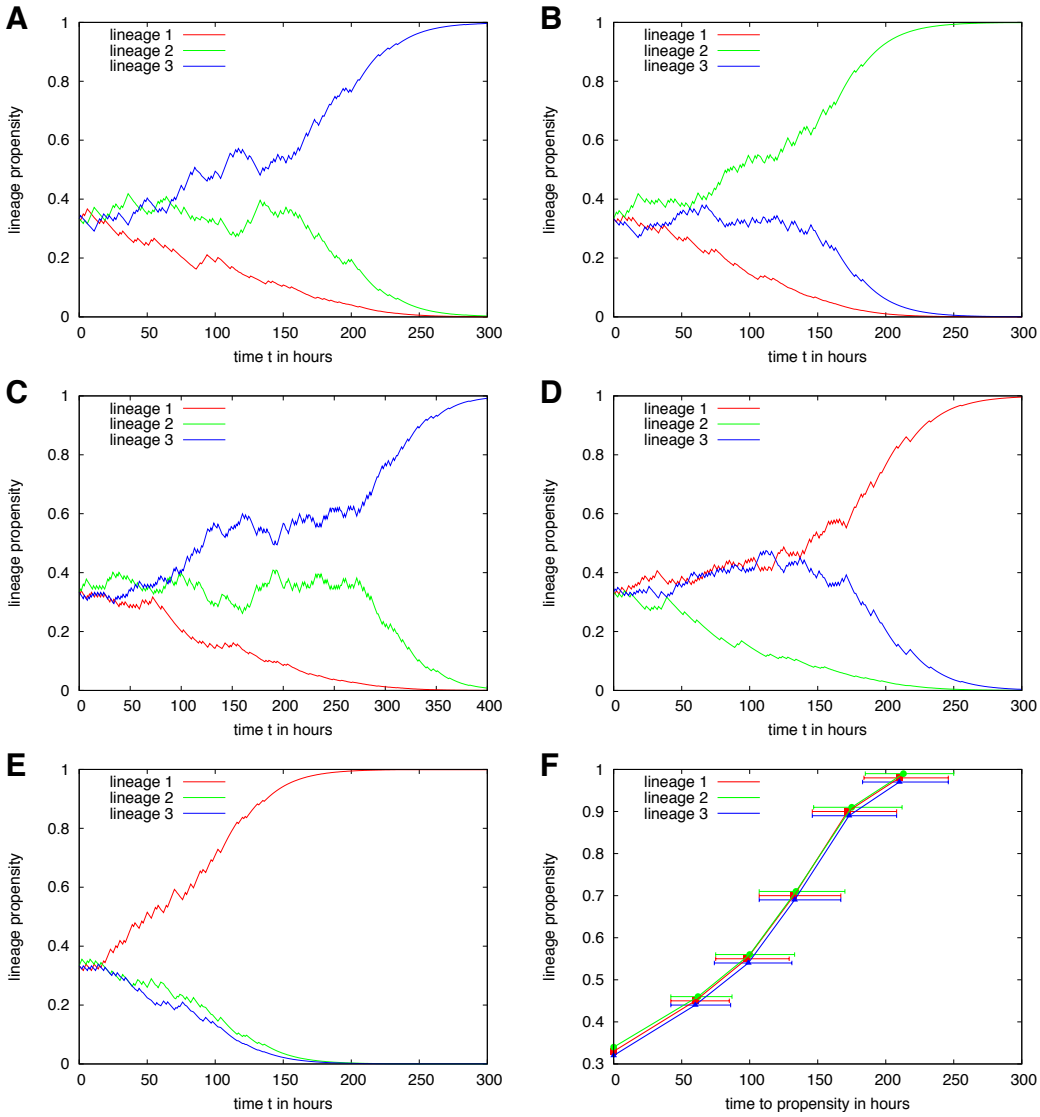

Figure 5.5.: Trajectories in the dissipative control regime.
(A - E) show a series of five possible realizations of the lineage specification process under the dissipative control regime (lineage propensity $\mathbf{x}$ vs. time for $N = 3$, $n_i = 0.05$, initialized with $\mathbf{x}(t = 0) = \{0.33, 0.33, 0.33\}$). **(F)** illustrates the characteristic divergence curve for the dominant lineage averaged over 10000 independent simulation runs. Starting from $x_i(t = 0) = 0.33$, threshold values (indicated by the squares) are set to $x^t = \{0.45, 0.55, 0.7, 0.9, 0.98\}$. Error bars indicate the interquartile range. Curves are slightly shifted within the figures to avoid superpositions.

lineage specification process. Finally, the normalization procedure slows down the lineage specification process and the leading lineage propensity $x_i^*(t) \to 1$ in the limit $t \to \infty$. Figures 5.5(A-E) show examples of five possible realizations with slightly different behavior during the initial competition phase. These differences are caused by the intrinsic stochasticity of the underlying update procedure. Further details of the simulation procedure are provided in Appendix D.2.

The set of realizations in Figure 5.5(A-E) suggests that the average divergence from the mean lineage propensity level $\bar{x} = 1/N$ is a suitable measure to characterize the dynamics in the dissipative control regime (*divergence curve*). Technically, for each individual trajectory the time point is recorded when the lineage propensity of the dominating lineage x_i^* passes a certain threshold x^t. Time averages for different individual threshold values are obtained separately for each dominating lineage i. The resulting divergence curves in Figure 5.5(F) connect the time averages for the individual threshold values $x^t = \{0.45, 0.55, 0.7, 0.9, 0.98\}$, thus providing a comprehensive picture of the temporal extension of lineage commitment for each individual lineage i.

Furthermore, the fraction of cells that finally commit to each of the N possible lineages is an indicator of the intrinsic balance of the lineage specification process. The importance of this measure is illustrated below.

Dependence on the reward. The critical parameter of the stochastic lineage specification process in the dissipative control regime is the intercept of the linear reward function f_m, defined as n_i in equation (4.5), which is constant and independent of x_i. As Figure 5.6(A) indicates, the process of lineage specification slows down as the reward $m_i = n_i$ declines with the intercept. However, as long as the intercepts are chosen to be identical for all lineages N also the contribution to each of the lineages is balanced. This is shown in Figure 5.6(B) outlining the fraction of cells that finally commit to each of the $N = 3$ possible lineages for different values of n_i.

Increasing *one* of the individual intercepts n_i results in a higher likelihood for the promotion of the particular lineage. Recalling that the intercept n_i directly determines the reward m_i it follows that the particular lineage with the highest reward benefits most in the update procedure. This "unbalanced" contribution is illustrated in Figure 5.7 in which the intercepts n_i are fixed for two of the three lineages ($i = 2, 3$, green and blue) whereas the other one is stepwise increased ($i = 1$, red). Figures 5.7(A), (C), and (E) illustrate the characteristic divergence curves. Given the increased reward of the "red" lineage the commitment process is accelerated as compared to the "blue" and "green" lineages.

Furthermore, the lineage contribution is skewed towards the preferred lineage. This behavior is illustrated by the bar charts in the corresponding Figures 5.7(B), (D), and (F), respectively, in which the fraction of cells in each of the $N = 3$

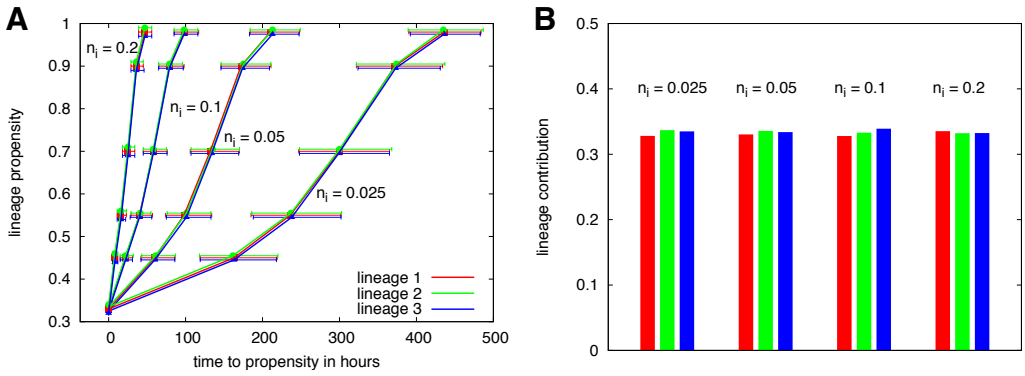

Figure 5.6.: Lineage commitment for balanced rewards.
(A) illustrates divergence curves (averaged over 10000 individual realizations with $\mathbf{x}(t=0) = \{0.33, 0.33, 0.33\}$) for different rewards $m_i = n_i = 0.025, 0.05, 0.1, 0.2$. For each of the four scenarios, n_i is identical for all $N = 3$ lineages ("balanced"). Bar charts in **(B)** show the fraction of cells finally committing to each of the lineages (indicated by corresponding colors) for the different scenarios $n_i = 0.025, 0.05, 0.1, 0.2$. Error bars are negligible due to many realizations.

possible lineages is shown. It is evident that the skewing effect becomes more dominant for increased intercept levels n_i of the preferred lineage.

In Figure 5.8 a similar scenario is shown in which only the intercept n_i of the "blue" lineage ($i = 3$) is kept fixed. The intercepts of the "red" and the "green" lineages ($i = 1, 2$, respectively) are both increased (in Figures 5.8(A-D): $n_1 > n_2 > n_3$, in Figures 5.8(E), (F): $n_1 = n_2 > n_3$). As indicated in the corresponding bar charts these scenarios lead to different skewed lineage contributions with an increasing disadvantage of the "blue" lineage ($i = 3$).

Dependence on number of lineages. Unlike the regressive control regime in which the number of competing lineages has little impact on dynamics of convergence towards the mean coexpression level $\bar{x}$, this effect is more pronounced in the dissipative regime. Upon initiation of the competition process in the dissipative regime there are in general N different lineages with similar probabilities to be chosen in the update process. As the individual update probability of a particular lineage decreases with the number of competitors it is evident that the time point until one of the lineages dominates over the others is delayed. This effect is illustrated in Figure 5.9(A) in which the divergence curves are compared for the scenarios $N = 3, 5, 7$. Figures 5.9(B), (C), (D) provide sample trajectories for the cases $N = 3$ (B), $N = 5$ (C), and $N = 7$ (D). These individual realizations support the notion that an increase in the number of lineages leads to a prolonged competition phase in which most of the lineages have a similar propensity levels.

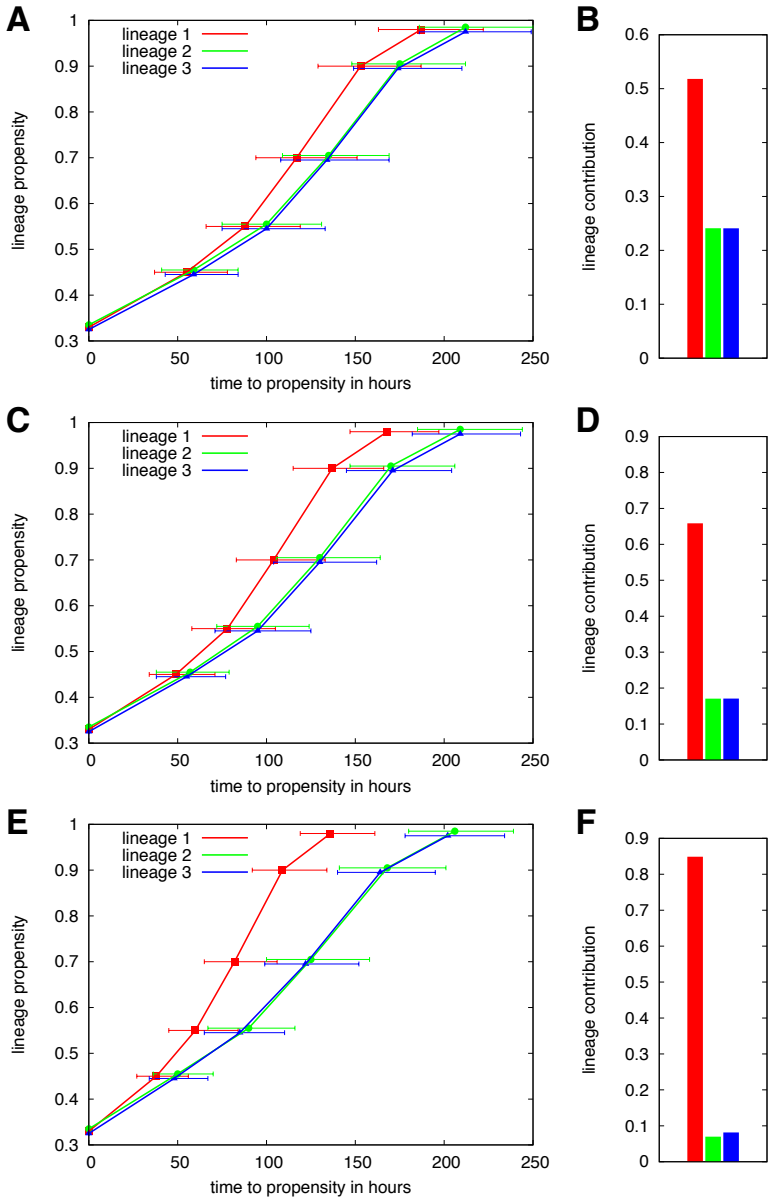

Figure 5.7.: Lineage commitment for unbalanced rewards (1).
(A), (C), (E) illustrate changes in the divergence curves for different unbalanced rewards. Intercepts of the lineages $i = 2$ (green) and $i = 3$ (blue) are kept fixed ($n_2 = n_3 = 0.05$) and the intercept of lineage $i = 1$ (red) is subsequently increased ($n_1 = 0.055$ in (A), $n_1 = 0.06$ in (C), $n_1 = 0.07$ in (E)). Averages for each scenario are based on 10000 individual realizations initialized with $\mathbf{x}(t = 0) = \{0.33, 0.33, 0.33\}$. The bar charts in **(B), (D), (F)** show the corresponding fractions of cells finally committing to each of the 3 lineages (lineage contribution, indicated by the color coding).

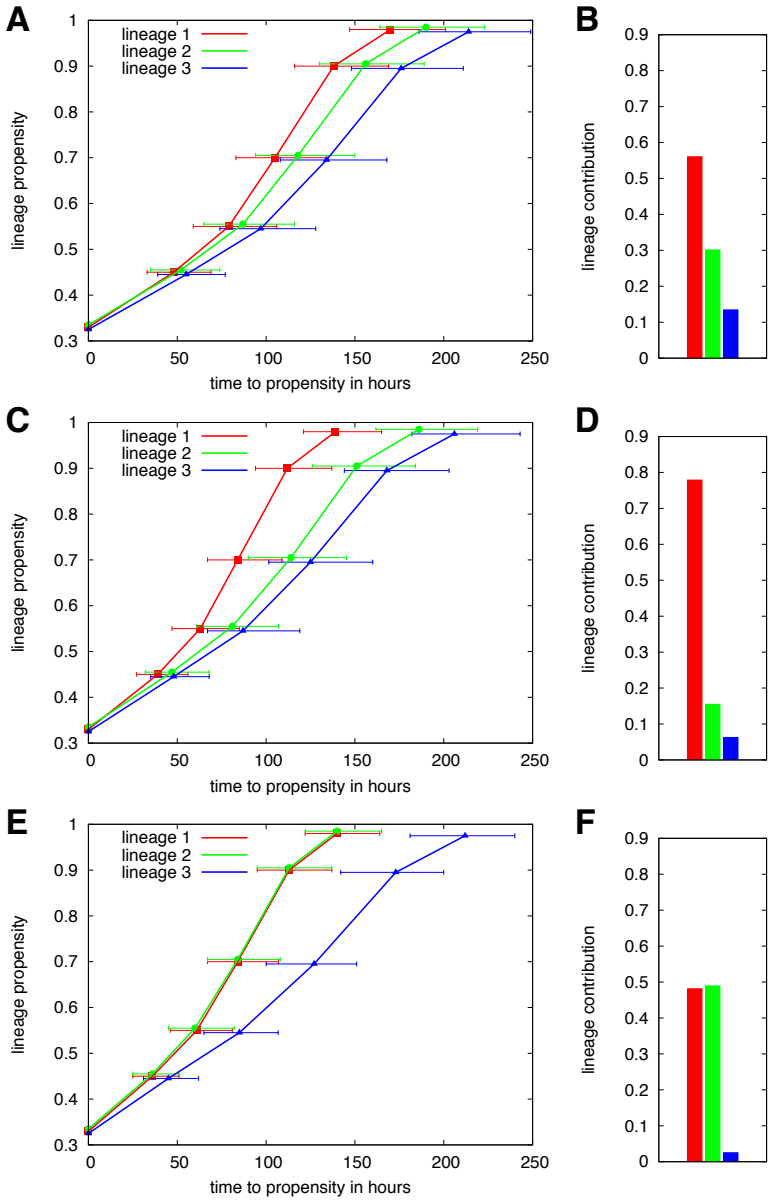

Figure 5.8.: Lineage commitment for unbalanced rewards (2).
(A), (C), (E) illustrate changes of the divergence curves for different unbalanced rewards. The intercepts of lineage $i = 3$ (blue) is kept fixed at $n_3 = 0.05$ for all scenarios. However, the intercepts of lineages $i = 1$ (red) and $i = 2$ (green) are subsequently increased ($n_1 = 0.06$, $n_2 = 0.055$ in (A), $n_1 = 0.07$, $n_2 = 0.055$ in (C), $n_1 = 0.07$, $n_2 = 0.07$ in (E)). Each scenario represents an average over 10000 individual realizations initialized with $\mathbf{x}(t = 0) = \{0.33, 0.33, 0.33\}$. The bar charts in **(B), (D), (F)** show the corresponding fraction of cells finally committing to each of the 3 lineages (lineage contribution).

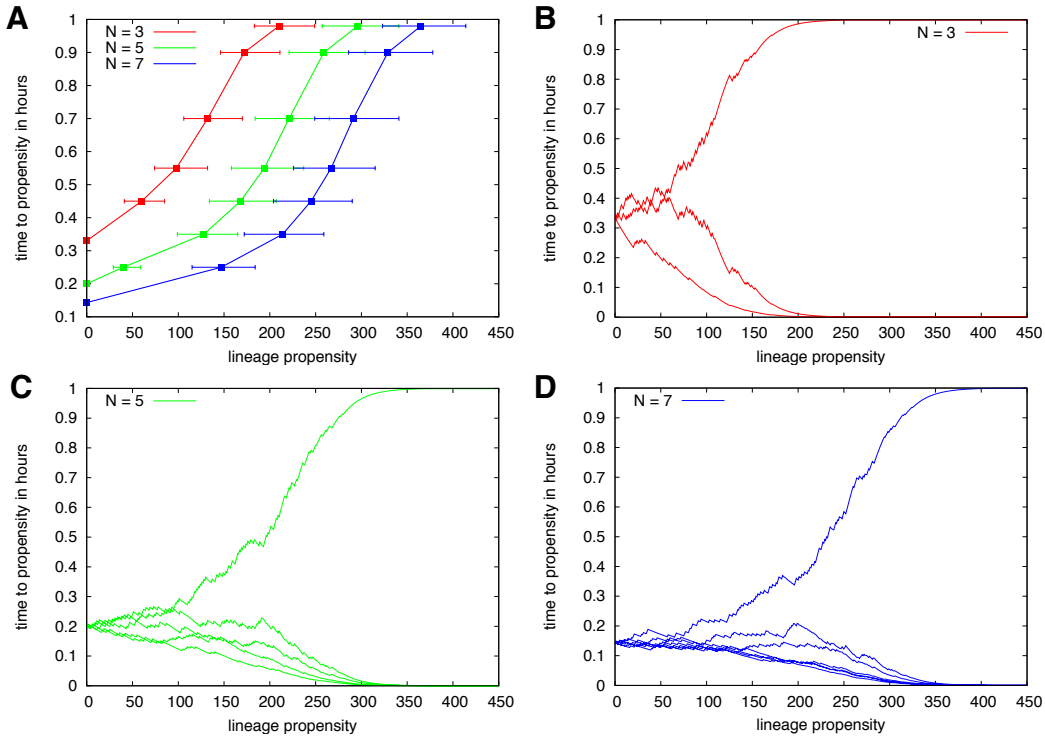

Figure 5.9.: Lineage commitment depending on the number of lineages. **(A)** shows the divergence curves for systems with $N = 3, 5, 7$ competing lineages. Averages are taken over 10000 individual realizations, each initiated with identical propensities at time $t = 0$, i.e. $x_i = 1/N$ for all i. The intercept of the linear reward functions has been set to $n_i = 0.05$, $\forall i$. **(B - D)** illustrate typical, individual realizations for the three scenarios $N = 3, 5, 7$, respectively.

Reversibility. Within the dissipative control regime it is a central question until which point of development the process of lineage commitment is still reversible. This issue is best addressed by analyzing the propensities of the lineages that are *not* dominating at the end of the commitment process (referred to as *inferior* lineages of a particular realization). The maximum value of the lineage propensity ($\max x_j$) for any inferior lineage j is an indicator for how far the dissipative process can continue while it is still possible to revert to another dominating lineage i which is finally up-regulated ($x_i^* \to 1$).

Figure 5.10 provides an estimate of this measure for different values of the reward $m_i = n_i$. For higher values of m_i there are a few cases for which the lineage propensity for an inferior lineage was already up to $x_j \approx 0.7$ but the process still reverted to another dominating lineage. As for smaller values of the reward $m_i = n_i$ more and more simulation steps are necessary to actually realize such a reversion these maximal propensity for the inferior lineages decrease progressively.

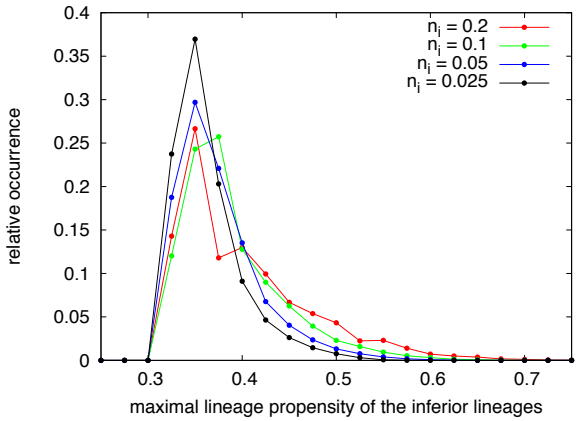

Figure 5.10.: Reversibility of the commitment process.
The figure shows the relative frequency of observing a maximal lineage propensity of the inferior lineages before the process of lineage commitment finally reverts to an alternative dominating lineage. Inferior lineages are all the lineages $j \in N$ that do not dominate a particular realization of a commitment process for $t \to \infty$. Different colors correspond to different values of the reward $m_i = n_i = 0.025, 0.05, 0.1, 0.2$. The points indicate relative frequencies of a corresponding histogram, connected by solid lines for better visualization. Histograms are taken from 500,000 realizations each, initialized with $\mathbf{x}(t = 0) = \{0.33, 0.33, 0.33\}$.

5.2. Dynamics of the extended stem cell model

5.2.1. Modeling the *in vivo* situation

As proposed in Section 4.4 the individual dynamics of lineage specification in the regressive and the dissipative control regime are embedded in the model of hematopoietic stem cell organization originally proposed by Roeder and Loeffler [33]. The overall dynamics of cell switches between signaling contexts $\mathcal{A}$ and Ω is not influenced by the dynamics of lineage specification. However, the individual, cell-intrinsic lineage propensity vectors $\mathbf{x}$ are modulated according to the rules of the regressive control regime for cells in $\mathcal{A}$ and according to the rules of the dissipative regime for cells in Ω. Although little information is available about the variability of transcriptional activity during the cell cycle, the model assumes without loss of generality that the process of lineage specification in signal context Ω only takes place in G1 but not during S, G2 and M phase.

Following the process of lineage specification within a single cell and its progeny over extended time periods, sample trajectories now appear as a superposition of the two different control regimes. The background color for the four sample trajectories shown in Figure 5.11 encodes the signaling context (and thus the control

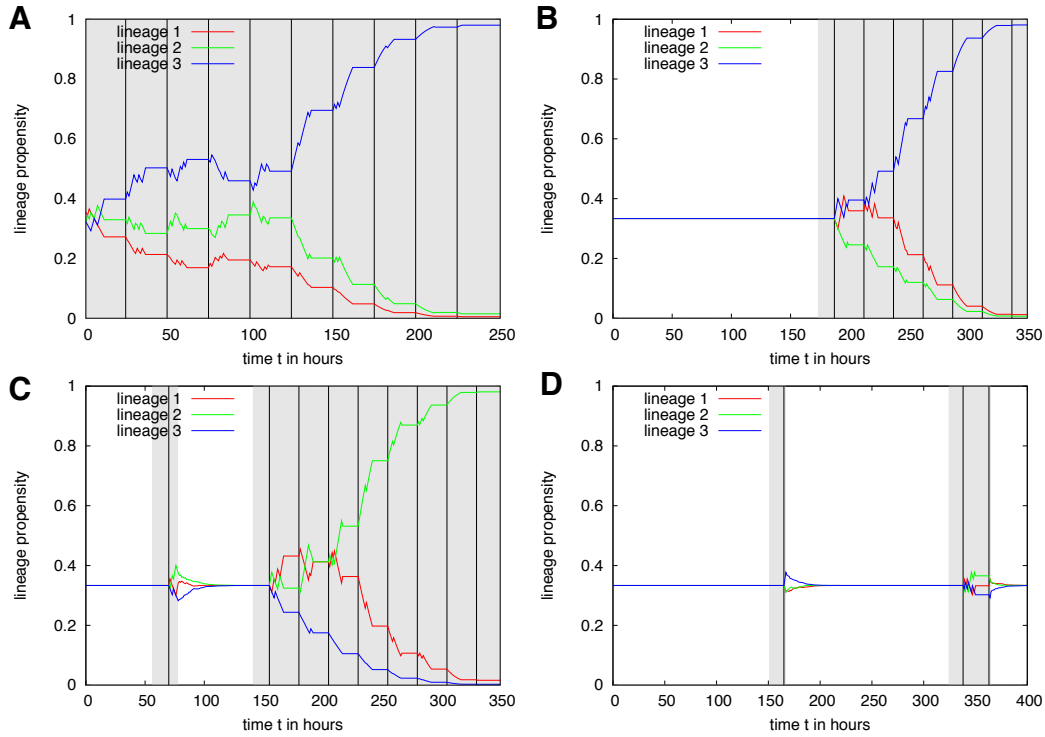

Figure 5.11.: Trajectories of the extended stem cell model.
(A - D) Possible realization of the lineage specification process within four sample cells chosen from one system (lineage propensity $\mathbf{x}$ as a function of time for $N = 3$ lineages, indicated by the colors). Upon cell division (thin black lines), lineage propensities are only followed in one of the two daughter cells. Background colors indicate the current signaling context of the cells (white background - $\mathcal{A}$, grey background - Ω). Parameters are set to $n_i = 0.1$ for the reward function in the dissipative regime, and $\sigma_i = 0.1$, $b_i = -1$, and $x_i^R = 1/3$ for $i = 1, 2, 3$ for the sigmoid reward function in the regressive control regime.

regime) to which the cell currently belongs. Whereas the white background indicates signaling context $\mathcal{A}$ (with the regressive control regime), signaling context Ω (with the dissipative control regime) is encoded by the grey background. The short intermediate plateaus in Ω correspond to the S, G2 and M phase of the cell cycle in which the process of lineage specification is halted. Technically, the time courses in Figure 5.11 show the continuation of the lineage specification process over several subsequent cell generations, thus only following the lineage propensities $\mathbf{x}$ in one of the two daughter cells after each cell division (black horizontal lines).

Initializing the complete model with a sufficient number of stem cells the system reaches a dynamically stabilized equilibrium with respect to the stem cell numbers

in $\mathcal{A}$ and Ω as well as the numbers of differentiating cells[1](see Figure 5.12(A)). Appropriate parameter configurations are provided in Appendix D.3. On top of this cellular dynamic the cell-intrinsic process of lineage specification leads to the establishment of stable fractions of cells committed to each of the N possible lineages. As for the example shown in Figure 5.12(B), $N = 3$ different lineages compete with identical parameters of the lineage specification process (dissipative regime $n_i = 0.1$, regressive regime with sigmoid reward function with saturation level $\sigma_i = 0.1$, slope $b_i = -1$). As the time courses indicate the three levels are fluctuating around the same mean level. The fluctuations are due to the limited life time of the cells, after which large clones of cells are deleted at identical times. Figure 5.12(C) shows the average fraction of cells committed to each of the three lineages as well as the fraction of uncommitted cells measured from time step 1000 to 4000.

As illustrated in the previous section, the intercept n_i of the linear reward function $f_m = m_i = n_i$ in the dissipative control regime is the critical parameter for the regulation of lineage specification. Different representative configurations of reward parameter $m_i = n_i$ are discussed below:

- The sequence of graphs in Figure 5.12 compares two model systems with different rewards $m_i = n_i$ in the dissipative control regime, namely the rewards for all lineages i are increased from $m_i = n_i = 0.1$ (Figures 5.12(A-C)) to $m_i = n_i = 0.2$ (Figures 5.12(D-E)). As expected, the overall cellular dynamic (Figures 5.12(A) and (D)) is left untouched as the parameters of the underlying HSC model (compare Section 3.4) remain constant. Similarly, the proportion of cells committing to each of the three lineages remains balanced as the probabilities for commitment are still equal (Figures 5.12(C) and (F)). In contrast, the overall increase of the rewards ($m_i = n_i = 0.1 \rightarrow m_i = n_i = 0.2$) accelerates the process of lineage specification. This reduces the fraction of uncommitted cells in which the dominant lineage propensity is below the threshold of commitment $x_i^* < x_{\text{com}}$ (shown in grey in the Figures 5.12 (B), (C) and (E), (F)).

- In the unbalanced situation, the individual rewards $m_i = n_i$ for the different lineages i differ. Figures 5.13(A-C) show a simulation sequence, in which two of the rewards are left untouched ($m_{1/2} = n_{1/2} = 0.1$) and one is slightly reduced ($m_3 = n_3 = 0.09$). Again, the overall cellular dynamic is left untouched (Figure 5.13(A)). In contrast, the probability of promoting the third lineage ($i = 3$, blue) is reduced. This effect is visible in the time courses (Figure 5.13(B)) as well as in the averaged fractions (Figure 5.13(C)).

[1]Referring to Figure 4.7 the *stem cell numbers in* $\mathcal{A}$ comprise all cells within signaling context $\mathcal{A}$, *stem cell numbers in* Ω refers to all cells in signaling context Ω with $a > a_{\min}$ and *numbers of differentiating cells* includes all cells in signaling context Ω with $a < a_{\min}$.

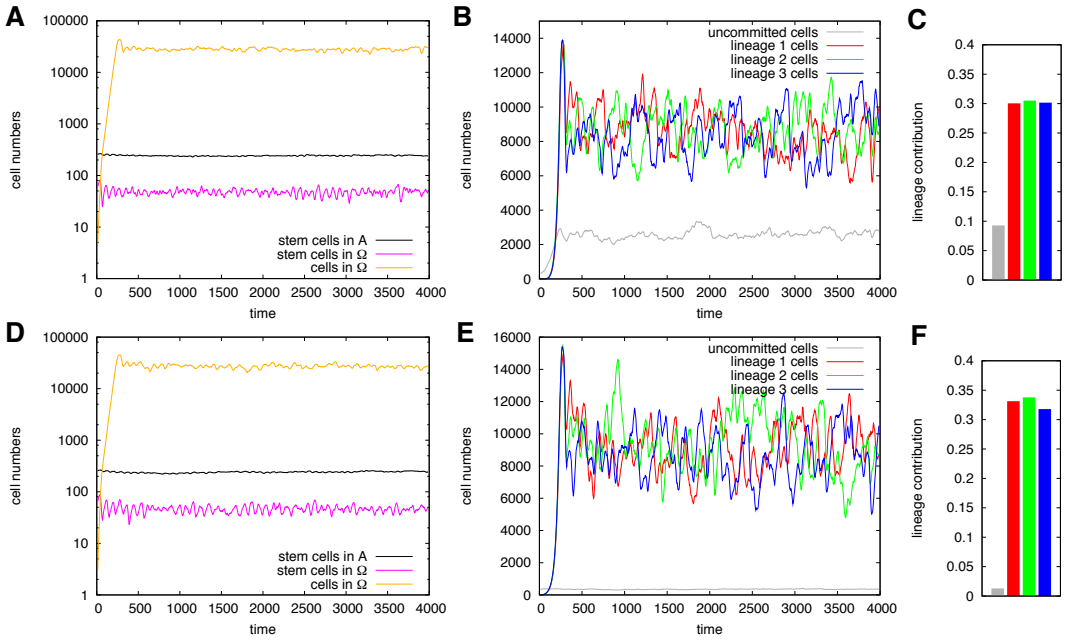

Figure 5.12.: Dynamics in system with balanced rewards.
(A), (D) show the number of stem cells in $\mathcal{A}$ (black), stem cells (magenta) and differentiating cells (orange) in Ω as functions of time starting from a sparsely populated initial system. **(B), (E)** show the corresponding time courses for the same system with respect to lineage commitment (grey - uncommitted cells; red, blue, green - cell numbers of the committed cells with $x_i^* > x_{\text{com}} = 0.75$). Bar-charts in **(C), (F)** indicate the fraction of cells in each of the four populations, averaged over the simulation run between time steps 1000 and 4000. Parameters of lineage specification are chosen as $n_i = 0.1$ (A - C) and $n_i = 0.2$ (D - F) for the reward function in the dissipative regime, and $\sigma_i = 0.1$, $b_i = -1$, and $x_i^R = 1/3$ for $i = 1, 2, 3$ for the sigmoid reward function in the regressive control regime.

- This effect is intensified if all lineages are parameterized with individual intercepts n_i. This is the case in the sequence of graphs in Figures 5.13(D-F) in which the intercepts are chosen as $n_1 = 0.12$, $n_2 = 0.1$, $n_3 = 0.08$. Clearly, lineage $i = 1$ (red) dominates the commitment process.

- Compared to the sequence of graphs in Figures 5.13(D-F), the simulations for Figures 5.13(E-J) are based on the doubling of the individual intercepts n_i for the three lineages ($n_1 = 0.24$, $n_2 = 0.2$, $n_3 = 0.16$) keeping their ratio constant. Although the commitment process is generally accelerated the overall appearance (one lineage high, one intermediate, one low) is maintained.

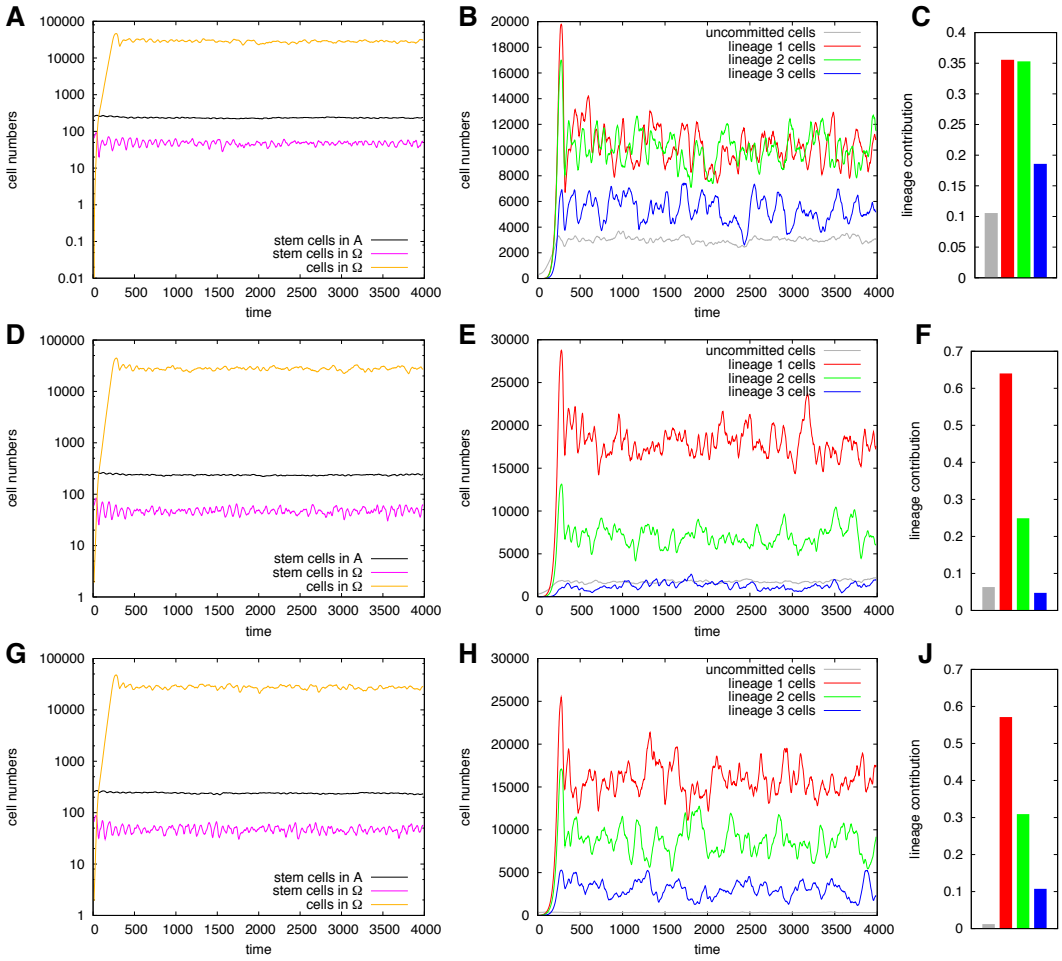

Figure 5.13.: Dynamics in a system with unbalanced rewards.
Time courses of the cellular composition and the number of committed cells in each lineage as well as the averaged proportions of cells are arranged similar to Figure 5.12. Parameters of lineage specification in the dissipative regime are chosen as $n_1 = n_2 = 0.1$ and $n_3 = 0.09$ (A-C); $n_1 = 0.12$, $n_2 = 0.1$ and $n_3 = 0.08$ (D-F); and $n_1 = 0.24$, $n_2 = 0.2$ and $n_3 = 0.16$ (G-J).The sigmoid reward function in the regressive regime is given by $\sigma_i = 0.1$, $b_i = -1$, and $x_i^R = 1/3$ for $i = 1, 2, 3$. Cells are counted as committed if $x_i^* > x_{\text{com}} = 0.75$.

Taking these results together, it can be demonstrated that the absolute rewards $m_i = n_i$ of the linear reward function in the dissipative control regime determine the overall "speed" of lineage commitment [2]. However, it is the ratio between the

[2]The "speed" of lineage commitment is independent form the "speed" of differentiation. The later is defined by the differentiation rate d that has been introduced in Section 3.4

individual lineage specific values that determine whether the commitment process is balanced or not.

Interestingly, the model system is less sensitive to the defining parameters of the regressive control regime. Even in the situation that the parameters of the reward functions in the regressive control regime differ considerably, there is almost no visible effect on the cellular level (data not shown). Recalling that the regressive regimes established the coexpression scenario in which all lineages propensities $\mathbf{x}$ are equalized, the corresponding parameters only alter the "speed" of this process. Differences in the lineage specific, individual parameters do not accumulate as long as the convergence to the mean expression levels $\bar{x}$ is generally shorter as the average residence time in $\mathcal{A}$.

The individual trajectories in Figure 5.11 indicate that cells changing back from signaling context Ω into $\mathcal{A}$ are involved in a process of fate reversion. As a certain time in Ω promotes lineage specification under the dissipative control regime, the cell's lineage propensity levels start to diverge from the common coexpression level. However, as the model explicitly relies on a dynamically stabilized equilibrium of cells in $\mathcal{A}$ and Ω, the process of fate reversion under the regressive control regime appears as a natural consequence. Assuming that the process of lineage specification is slow enough compared to the speed of differentiation (loss of affinity a described by the model parameter d) this reversion only occurs for early steps of the lineage commitment. However, in more extreme scenarios, as in the case of transplantation experiments in irradiated recipients, this effect might be more pronounced.

Briefly summarizing, the model has been constructed such that the finally dominating lineage of a particular realization can only be predicted in a probabilistic sense. In the case of identical rewards n_i in the dissipative regime, the dominance is equally likely for each lineage i . However, the specification of rewards n_i with different values for each lineage skews the decision process towards lineages with the higher rewards. Moreover, a self-renewing cell can contribute to different cells types in varying fractions.

5.2.2. Adapting the model to the *in vitro* situation

As shown in a previous study [223], the experimental *in vitro* situation is sufficiently captured by reducing the regeneration rate r and manipulating the $a_{\min}$ limit. These assumptions generally lead to an exhaustion of the population of self-renewing cells and the initiation of lineage commitment processes within all cells. However, the dynamics of the lineage specification process remain largely untouched and the phenomena of reduced self-renewal capacity is merely attributed to the original cell-kinetic model.

Since no stabilized dynamic equilibrium is established the composition of the cell culture is generally accessed in a time-dependent manner or by end point analysis.

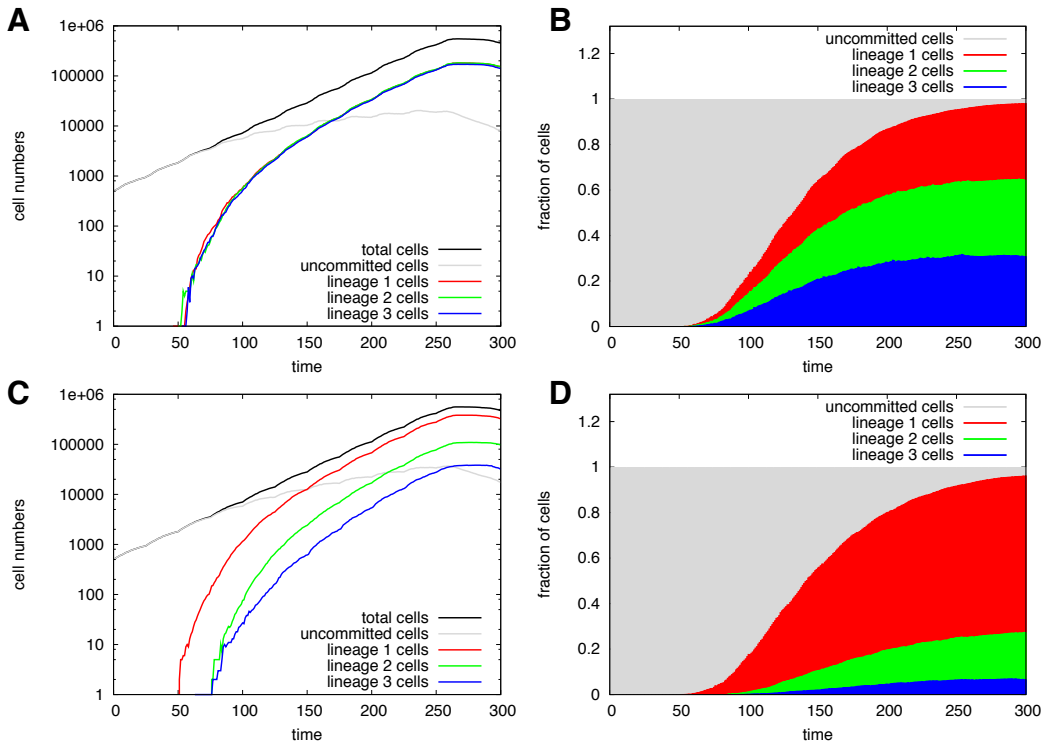

Figure 5.14.: Dynamics of an *in vitro* system with balanced and unbalanced rewards.
Time courses of the cellular composition are shown for a simulated *in vitro* scenario (regeneration rate $r = 1$, norm number of signaling context $\mathcal{A}N_A = 5$, initiated with 500 cells within $a \in [0.01, 1]$). **(A)** Absolute cell numbers as a function of time for a scenario with balanced contribution to all lineages (black - total cells, grey - uncommitted cells, colored - lineages 1 to 3). **(B)** Fraction of cells being uncommitted (grey) versus committed to lineages 1, 2 or 3 (red, green, blue) as a function of time for the same balanced rewards ($m_{1...3} = n_{1...3} = 0.1$). **(C), (D)** Corresponding graphs to (A) and (B) for the situation of unbalanced rewards ($m_1 = n_1 = 0.1$, $m_2 = n_2 = 0.08$, $m_3 = n_3 = 0.07$). The sigmoid reward function in the regressive regime is set to $\sigma_i = 0.1$, $b_i = -1$, and $x_i^R = 1/3$ for $i = 1, 2, 3$ in all scenarios. Cells are counted as committed if $x_i^* > x_{\text{com}} = 0.75$.

Figure 5.14 provides typical time courses simulating an exponentially expanding cell culture under *in vitro* conditions. As shown in Figure 5.14(A), the first cells appear as committed around time step $t = 50$. Following their expansions, the number of uncommitted cells (grey) diverges from the exponentially increasing total number of cells. Subsequently, the committed cells determine the overall system dynamics before the system finally declines due to the limited lifetime of the differentiating cells. Figure 5.14(B) provides almost identical information.

Here, the fraction of cells within each lineage is shown instead of the total cell numbers. This is conceptually closer to cell culture experiments in which the phenotypic composition is accessed over time. Figures 5.14(C) and (D) illustrate the same behavior for the situation of unbalanced lineage contribution in which the development of the "red" lineage is more likely as compared to the other lineages. As for the modeling of the *in vivo* situation, the skewed lineage contribution results by adjusting the intercepts in the dissipative regime (for the shown situation: $n_1 > n_2 > n_3$).

5.2.3. Instructive versus selective lineage specification

As pointed out in Section 4.3.4 of the previous chapter, there are in principle two scenarios to influence lineage contribution, namely the *instructive* and the *selective* scenario of lineage specification. The previous analysis in this section focused particularly on the question how lineage contribution can be intrinsically skewed by unbalanced rewards. This approach directly corresponds to the *instructive* lineage specification since subsequent cell death does not have a regulating function on the lineage contributions.

In contrast, *selective* lineage specification is based on the concept that the initial lineage decision is balanced (by using identical rewards m_i for all lineages) whereas the regulation of the lineage contribution is facilitated by a subsequent, selective cell death process. Details of the selection process are described in Section 4.3.4.

In order to illustrate the dynamics of *selective* lineage specification, a system is studied in which the contributions to each of $N = 3$ possible lineages is balanced on the level of the intrinsic lineage decision by using identical rewards m_i for all lineages i. As a reference system, the selective cell death process Φ^S is adjusted such that the lineage specific intensities for cell death (defined in equation (4.9)) are set to $\phi_i = 0.01$ for all lineages i (also, the boundaries for the action of the cell death process $x_i^{\text{death/low}} = 0.5$ and $x_i^{\text{death/high}} = 1.0$ are set to identical values for all i). In this case, there is no selective effect as all lineages are equally affected by the cell death process Φ^S and the fraction of cells contributing to each of the possible lineages is conserved (Figure 5.15(A), (B)). Additional to the cell numbers in each lineage in Figure 5.15(A) also the cumulative number of death cells is shown (in yellow). As the cell death intensity $\Phi^S = \phi_i$ remains constant once a certain threshold for the dominating propensity is passed ($x_i^* > x_i^{\text{death/low}}$) the cumulative number of death cells follows the development of the target cell population.

In contrast, if the probability for cell death ϕ_i is increased for one lineage as compared to the others, the relative fraction of cells committed to this lineage decreases. This intuitive result is illustrated in Figures 5.15(C), (D). Going one step further and assigning different cell death intensities ϕ_i to each individual lineage, a qualitatively similar result is obtained as in the instructive scenario (compare Figure 5.14(C), (D)): one lineage is positively selected whereas the other

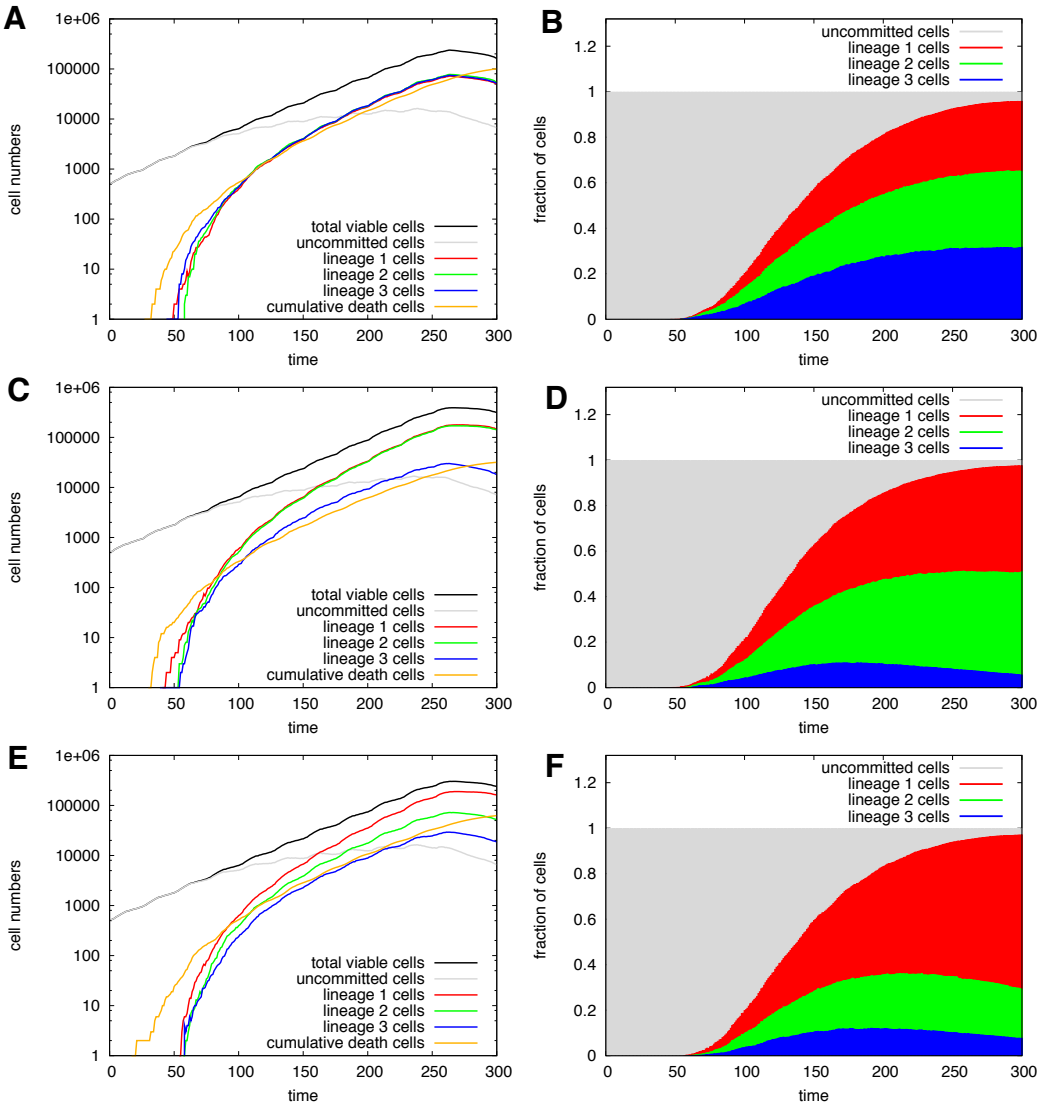

Figure 5.15.: Dynamics of an *in vitro* system with *selective* lineage specification.

(A), (C), (F) Time courses of the cellular composition including the cumulative number of death cells and **(B), (D), (E)** the fraction of cells are shown as functions of time for a simulated *in vitro scenario* with *selective* lineage specification (regeneration rate $r = 1$, norm number of signaling context $\mathcal{A}\,N_A = 5$, initiated with 500 cells within $a \in [0.01, 1]$). Colors correspond to Figure 5.14. Additionally, the number of death cells is shown in yellow. The scenarios differ in the lineage specific cell death intensity: (A), (B): $\phi_i = 0.01$ for $i = 1, 2, 3$; (C), (D): $\phi_1 = \phi_2 = 0.0$ and $\phi_3 = 0.02$; (E), (F): $\phi_1 = 0.0$, $\phi_2 = 0.01$ and $\phi_3 = 0.02$. Boundaries for the cell death action are set to $x_i^{\text{death/low}} = 0.5$ and $x_i^{\text{death/high}} = 1.0$. Parameters of the intrinsic lineage specification are identical for all lineages: in the dissipative regime $m_{1\ldots3} = n_{1\ldots3} = 0.1$ and for the sigmoid reward function in the regressive regime $\sigma_i = 0.1$, $b_i = -1$, and $x_i^R = 1/3$. Cells are counted as committed if $x_i^* > x_{\text{com}} = 0.75$.

lineages play only minor roles. A particular example for the selective scenario is shown in Figures 5.15(E),(F).

For reasons of completeness, it is remarked that selective lineage specification can also be applied for the simulations of the *in vivo* scenarios yielding stable populations of committed cells over time. The qualitative result, that some lineages can be promoted at the expense of others, holds in this situation, too. A comparative study between the selective and the instructive scenarios is provided in Section 5.4.3.

5.3. Fluctuations of the ground state and lineage bias

As mentioned above, a divergence of the root of the reward functions x^R from the mean propensity level $\bar{x} = 1/N$ induces the individual lineage propensities to fluctuate around their mean expression level instead of converging to this level. In order to understand this behavior one might follow a brief experiment of thought: Given that the root of the reward functions is higher than the mean propensity level ($x^R > \bar{x} = 1/N$) and assuming that N lineages have lineage propensities $x_i \approx 1/N$, the update procedure would choose each lineage with almost identical probability. This lineage would receive a positive reward (as $x_i \approx 1/N < x^R$), thus increasing its own propensity and decreasing the propensities of the others due to the normalization routine. However, at a certain point, an individual propensity x_i might exceed the root of the reward functions x^R, thus in the next update, for which this lineage has been chosen, it receives a negative reward. This imposes an upper limitation and keeps the whole update process in the regressive regime in a *frustrated* state: the system can neither diverge nor converge to a fixed level.

A similar explanation holds for the situation, that the root of the reward functions is smaller than the mean propensity level ($x^R < \bar{x} = 1/N$). In this case, the propensities preferentially decrease until they are limited by a lower boundary as soon as individual propensity levels decrease below x^R. The phenomenological behavior is illustrated in Figure 5.16 in which individual realizations are shown for the cases $x^R < \bar{x} = 1/N$ (Figure 5.16(A), (B)), $x^R = \bar{x} = 1/N$ (Figure 5.16(C), this is the general case in which the propensities converge towards $\bar{x} = 1/N$) and $x^R > \bar{x} = 1/N$ (Figure 5.16(D), (E)). Figure 5.16(F) shows the corresponding sigmoid reward functions which are just shifted with respect to their root x^R. Colors correspond to the individual realizations in Figures 5.16(A-E).

As long as the deviations between the root of the reward functions x^R and the mean propensity level $\bar{x} = 1/N$ are still moderate, all propensities can fluctuate around their common mean level. However, increasing the root of the reward functions such that the $x^R \to 1/(N-1)$ the system can reach a qualitatively different dynamical equilibrium. As indicated for three example realizations in Figure 5.17 with $x^R = 0.47, 0.5, 0.55$ the propensities in the case $N = 3$ self-

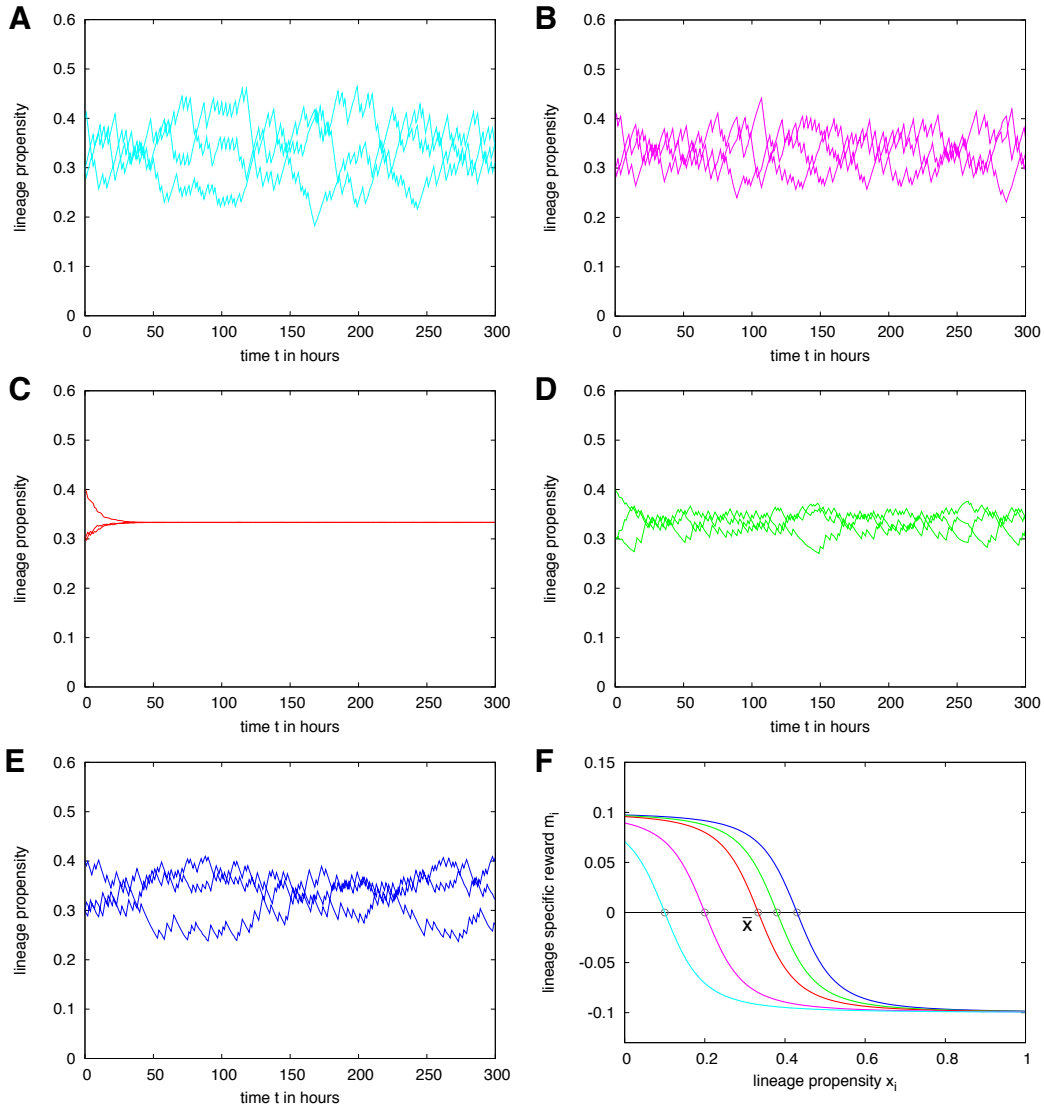

Figure 5.16.: Fluctuations of the individual lineage propensities for different values of the root x^R.
(A - E) Lineage propensities $\mathbf{x}$ for a system with $N = 3$ lineages are shown as functions of time for different values of the root of the sigmoid reward function $x^R = 0.1$ (A), 0.2 (B), 0.33 (C), 0.38 (D), 0.43 (E) in the regressive control regime. The saturation level $\sigma_{1...3} = 0.1$ and the slope $b_{1...3} = -1$ of the reward function are identical for all realizations. The shape of the corresponding reward functions is shown in Subfigure **(F)**. The coloring corresponds to the individual realizations in (A - E).

organize such that two of them are up-regulated while one is already suppressed. This phenomena is termed *lineage bias*, as upon switching to the dissipative regime the two up-regulated lineages are treated preferentially.

At this point a brief remark is in place to clarify definitions. The process of shifting lineage contributions from the balanced to the unbalanced situation (by tuning the rewards $m_i = n_i$) as discussed in Section 5.1.2 is referred to as *skewing* throughout this thesis. In these cases lineage propensities converge to $x_i = \bar{x} = 1/N$ in the regressive regime which corresponds to the parameter choise for the root value x^R. In contrast, the preference for certain lineages induced by a frustrated system with $x^R \neq \bar{x} = 1/N$ as outlined above is referred to as *lineage bias*.

For a system with $N = 3$ lineages and $x^R \leq 1/2 = 1/(N-1)$ (Figure 5.17(A); red sigmoid reward function in Figure 5.17(D)) the higher propensities stabilize at a level below $1/(N-1) = 1/2$ resulting in a non-zero value for the third, lower propensity. In contrast for the case that $x^R = 1/(N-1) = 1/2$ (Figure 5.17(B); green sigmoid reward function in Figure 5.17(D)) the system reaches a stable state with two propensities at $x_i = 1/(N-1) = 1/2$ and the third propensity approaching zero. This behavior is also observed for the case that $x^R \geq 1/(N-1) = 1/2$ (Figure 5.17(C); blue sigmoid reward function in Figure 5.17(D)). However, as $x^R > 1/(N-1)$ the two up-regulated propensities fluctuate around the level $x_i = 1/(N-1) = 1/2$.

Considering a large population of cells, the introduction of a lineage bias does not ultimately result in a shift in the lineage contributions. If the reward functions in the regressive control regime have identical parameters, the lineage propensities, which are up- or down-regulated during the establishment of the lineage bias, are equally distributed. Only in the case, that one lineage i is already penalized in the regressive regime (e.g. by choosing a smaller saturation level σ_i) it is less present among the up-regulated lineage propensities during the establishment of the lineage bias. If the cells are then switched into the dissipative regime, the particular lineage does already have an initial disadvantage leading to reduced lineage contribution.

This scenario is exemplified in Figure 5.18. After keeping 10000 cells in the regressive control for 1000 timesteps (with the root of the reward functions set to $x^R = 0,47,\ 0.5,\ 0.55$, respectively), a random lineage bias establishes within each of these cells. Thus two propensities are up-regulated while one is down-regulated. However, if this process is unbiased ($\sigma_{1\ldots3} = 0.1$), all lineages are equally likely to be among the up-regulated ones. Changing the system to the dissipative regime, the lineage contributions of the three individual lineages are still balanced as shown in Figure 5.18(A) for three different values of the root of the reward functions $x^R = 0.47, 0.5, 0.55$. However, if one lineage (in this case the blue one) has a disadvantage during the establishment of the lineage bias in the regressive control regime ($\sigma_{1,2} = 0.1$, $\sigma_3 = 0.08$), it is less present among the up-

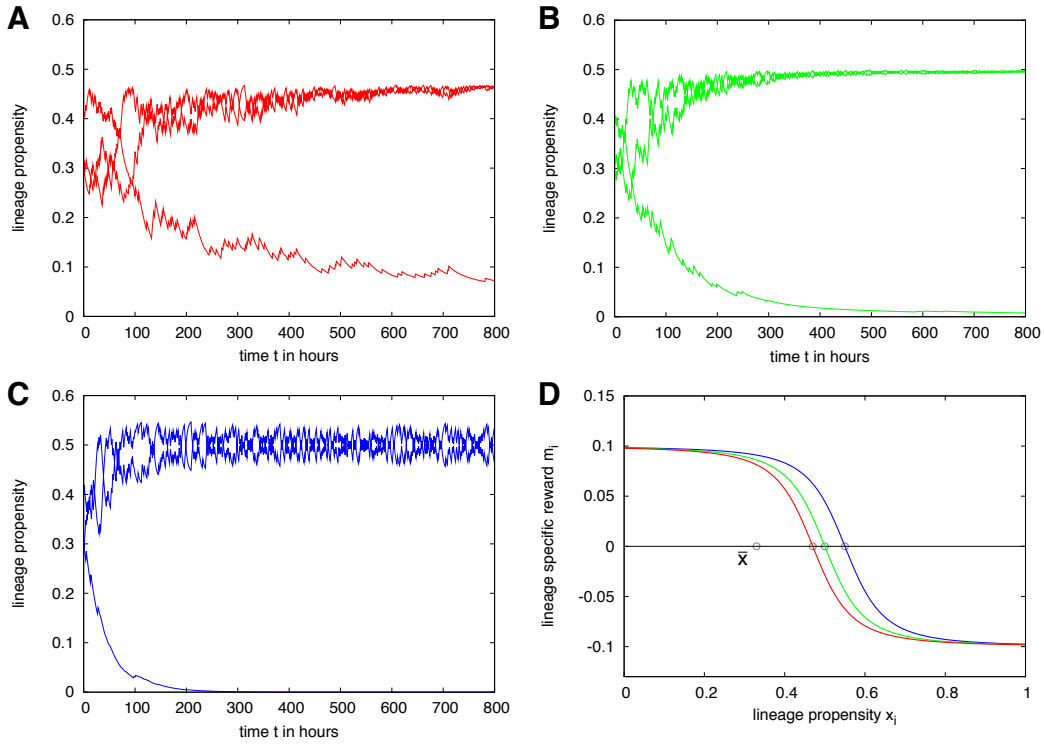

Figure 5.17.: Lineage bias for different values of the root x^R.
(A - C) Lineage propensities **x** for a system with $N = 3$ lineages are shown as functions of time for different values of the root of the sigmoid reward function $x^R = 0.47$ (A), 0.5 (B), 0.55 (C) in the regressive control regime. The saturation level $\sigma_{1...3} = 0.1$ and the slope $b_{1...3} = -1$ of the reward function are identical for all realizations. The shape of the corresponding reward functions is shown in **(D)**. The coloring corresponds to the individual realizations in (A - C).

regulated lineage propensities. After switching the cells to the dissipative control regime this effect becomes manifest by the reduced contributions of the "blue" cells (Figure 5.18(B)).

The analysis shows that the model is in principle able to represent an intrinsic lineage bias. A detailed comparison of this phenomenon to available data is beyond the scope of this thesis and will be subject to a separate scientific work.

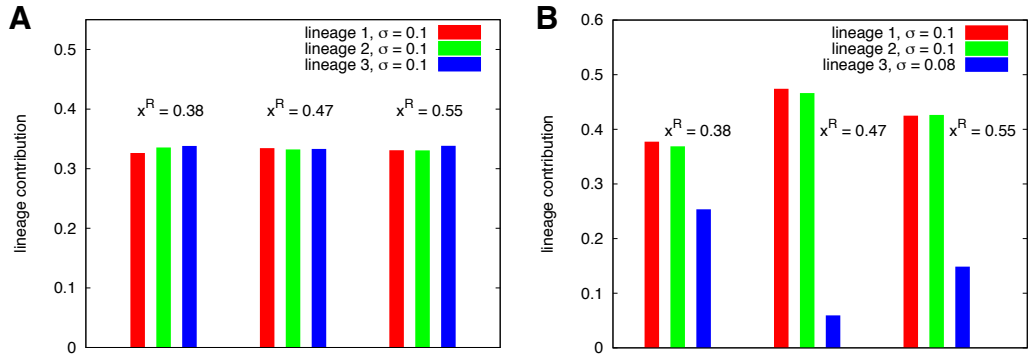

Figure 5.18.: Lineage contribution for the fluctuating ground state.
(A). 10000 cells are simulated in the regressive control regime with the root of the reward functions set to $x^R = 0, 47,\ 0.5,\ 0.55$. Further parameters are set to $x_{1...3}(t = 0) = 0.333$, saturation level $\sigma_{1...3} = 0.1$ and slope $b_{1...3} = -1$. For the simulations in **(B)** only the saturation level of the reward function of the third lineage $i = 3$ is reduced to $\sigma_3 = 0.08$. After 1000 time steps all cells are switched into the dissipative control regime with parameters $n_{1...3} = 0.05$. The contribution to the individual lineages is shown by the bar charts depending on the position of the root point in the regressive regime x^R.

5.4. Experimental validation

In the following the established simulation model of lineage specification is applied to a number of landmark experiments that characterize the developmental potential of hematopoietic stem and progenitor cells *in vitro*.

5.4.1. Lineage contribution of single differentiating cells

Early experiments on the lineage contribution of isolated hematopoietic, spleen-derived mouse cells revealed the existence of cells that give rise to more than one cell type [26, 27]. In a particular protocol, Ogawa and coworkers used a primary culture initiation assay to purify colony forming cells [26]. Within the population of colony forming cells, they found that about two thirds of the cells give rise to clonal colonies containing only one cell type, whereas the remaining third of the cells contributed to two or even more (up to six) different cell types. In a recent study, Takano et al. [222] used CD34- c-Kit+ Sca-1+ lin- (CD34- KSL) cells taken from adult mouse bone marrow for a similar set of experiments. The lineage composition of the progeny of these cells was analyzed *in vitro* using single cell differentiation assays.

For the comparison of the particular data with the proposed model, a simple simulation protocol is adapted. A previously established reference parameter

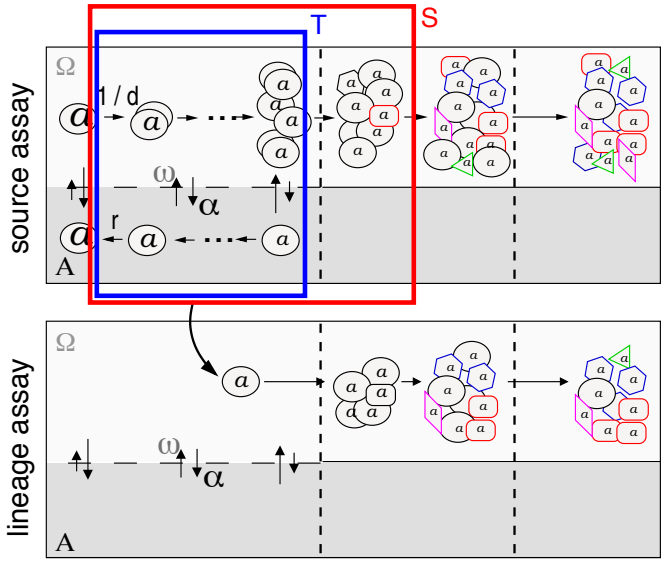

Figure 5.19.: Simulation strategies for single cell differentiation experiments.
Cells from a defined region within the homeostatic source assay (transfer pool) are transferred into a lineage assay. The transfer pools are indicated by the boxes in the source assay (pool S (red) - fit for the data from Suda et al. [26, 27]; pool T (blue) - fit for the data from Takano et al. [222]). Within the lineage assay the cells undergo lineage commitment and the contribution of different lineage types is evaluated.

set describing hematopoietic stem cell organization in unperturbed mice [221] is used for the representation of the homeostatic situation from which bone marrow cells had been isolated experimentally (*source assay*, compare Figure 5.19). This parameter set was complemented by an appropriate representation of the intracellular lineage specification dynamics with equal rewards m_i for all lineages, intentionally neglecting any correlations between the development of certain lineages. Parameters are chosen such that a differentiating stem cell is considered committed ($x_i^* > x_{\text{com}}$) after four to ten days of lineage specification in signaling context Ω. Cells used for transfer into the lineage assays are chosen randomly among a well defined subpopulation of the source assay (called *transfer pool*), characterized by the range of the affinity parameter a_{trans}. The boundaries of these transfer pools are the central parameters to fit the simulation results to the experimental data by Suda et al. [26, 27] (pool S, shown in red in Figure 5.19) and Takano et al. [222] (pool T, shown in blue in Figure 5.19). The *lineage assay* is represented by an empty model system in which the development of the progeny is observed for 240 h. Finally, the number and the lineage of cells produced in each

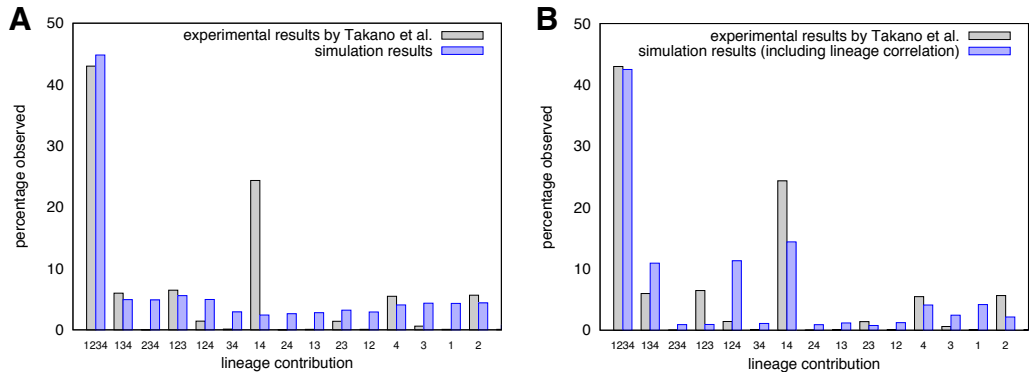

Figure 5.20.: Lineage contribution of single differentiating cells
(A). Experimental results for the differentiation of single CD34- KSL cells [222] are indicated by the grey bars (mean, CI not available). Results for the corresponding simulated cells are shown in blue (mean of 50,000 simulation runs). The lineage contribution is coded as follows: 1-neutrophil, 2-megakaryocyte, 3-erythroblast, 4-macrophage **(B).** Introduction of a moderate correlation between lineages 1 and 4 in the *in silico* model increases the particular lineage contributions. Parameters are provided in the text and in Appendix D.4.

lineage assay are evaluated. Due to an expected deficiency of a properly functioning hematopoietic niche environment in cell cultures it is assumed that for all simulated *in vitro* assays the signaling context $\mathcal{A}$ simply maintains the self-renewal ability of a cell (measured by its affinity a) but does not promote its regeneration ($r = 1$, see also Section 5.2.2).

In their series of experiments, Ogawa and coworkers [26, 27] distinguished six different lineages, namely neutrophils, macrophages, eosinophils, mast cells, megakaryocytes, and erythrocytes. Therefore, the *in silico* intracellular lineage specification dynamics are constructed with $N = 6$ different lineages. The transfer pool S (cf. Figure 5.19) for the simulation of the single cell differentiation data [26] has been adjusted to $a^S_{\text{trans}} \in [0.000001, 0.99]$. Using this parameterization the experimental observations are closely matched: two third of the cells only contribute to a single lineage (model prediction: 65.5 %) whereas the remaining third contributes to two or more lineages. The potential development in six different lineages yields 57 possible combinations in which the colonies contribute to two or more different cell types (multiple lineage contribution). However, as the authors report only about 50 such colonies in total, a statistical evaluation and a comparison with the simulation data is not instructive.

To adapt the model system to the experimental setup presented by Takano et al. [222], the number of possible lineages is reduced to N=4 (neutrophils, megakaryocytes, erythroblasts, macrophages). Due to the more sophisticated stem

cell sorting procedure used in this experiment, the source population of initial parent cells is expected to contain an increased fraction of uncommitted cells. This is reflected by the narrower transfer pool T ($a^T_{\text{trans}} \in [0.012, 0.99]$, cf. Figure 5.19) marking the difference to the simulations of the previous experiments by Suda et al. [26]. In the single cell differentiation studies of bone marrow derived CD34- KSL cells [222], the majority (43%) of plated cells contributed to all four determined lineages while other combinations are observed with lower frequency. Applying the adapted transfer pool T it appears that in the majority of cases (model result: 42.9%) the progeny contained all four lineages whereas other combinations are reduced (Figure 5.20(A)). It should be emphasized that this qualitative pattern is achieved even under the simplifying assumption of balanced rewards $m_i = n_i = m$. However, the precise matching of the results is incomplete. The experimental data suggests that there is a moderate correlation between neutrophil and macrophage differentiation (see lineage combination 1:4 in Figure 5.20(A)).

In order to study whether the artifical introduction of a moderate, positive correlation between two lineages results in this kind of behavior a minor modification of the presented model has been developed. The details are provided Section 4.3.2. For the particular case of a correlation between lineages 1 and 4 (neutrophils and macrophages, respectively) a positive correlation parameter $\gamma_{14} = \gamma_{41} = 0.3$ has been applied. For all other pairs of cell types i, j, these correlations are neglected, thus $\gamma_{i,j} = 0$.

As indicated in Figure 5.20(B), the *in silico* model leads to a shift in the differentiation pattern similar to the experimental observations. Progeny of single cells containing neutrophil and macrophage cells are now significantly enhanced compared to other developments. However, due to the complexity of these, potentially weak correlations between certain lineages and the limited amount of available date, a detailed quantification and further study of this process has not been performed.

Specific parameter settings for the simulations in this Section are provided in Appendix D.4.

5.4.2. Comparative differentiation of paired daughter cells

Extending the experiments on the differentiation of individual cells *in vitro*, Ogawa and coworkers examined the developmental fate of two daughter cells derived from one parent cell [27]. By comparing the lineage contribution within the colonies generated from each of the daughters, the authors concluded that lineage potential is progressively restricted by a sequence of stochastic commitment steps taking place at each cell division. However, as already discussed in Section 3.3.5, the coupling of the lineage restriction to the cell divisions has not been demonstrated convincingly. The proposed model of lineage specification, which is intrinsically decoupled form the process of cell division, further challenges this view.

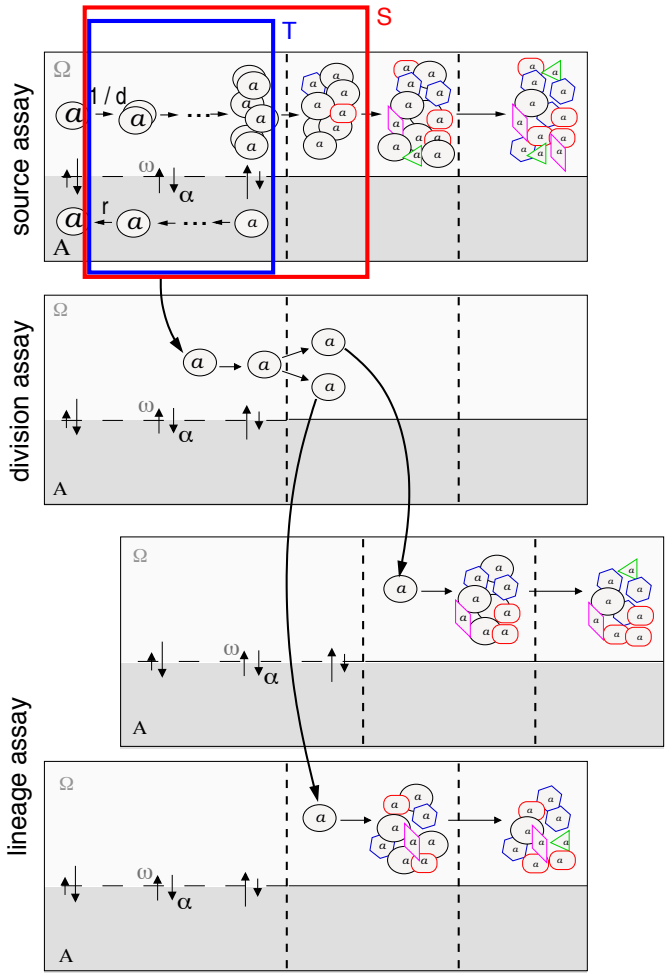

Figure 5.21.: Simulation strategies for paired daughter cell differentiation experiments.
Cells from a defined region within the homeostatic source assay (transfer pool) are transferred into a division assay. The transfer pools are indicated by the boxes in the source assay and are identical to the single cell differentiation experiments outlined in Figure 5.19 (pool S (red) - fit for the data from Suda et al. [27]; pool T (blue) - fit for the data from Takano et al. [222]). After the first division in the division assay both daughter cells are separated and their lineage contribution is evaluated within a lineage assay.

To support the analysis of this type of experiment in the context of the proposed model a further data set is taken into account. Takano and coworkers [222] repeated the earlier experiments by using a more sophisticated selection protocol to identify progenitor hematopoietic progenitor cells (CD34- KSL) taken from adult mouse bone marrow. The variability of the lineage contribution of a parental cell

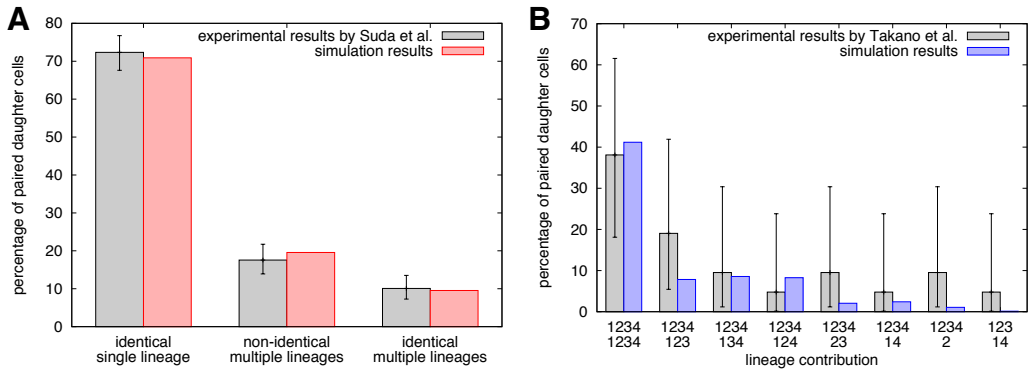

Figure 5.22.: Lineage contribution of paired daughter cells.
(A). Experimental results for hematopoietic spleen-derived mouse cells are shown in grey (mean, 95% CI) [27]. Results for the simulated cells (defined by the transfer pool S of the source assay) are shown in red (mean of 50,000 simulation runs, CI negligible due to the high number of replicates). **(B).** Experimental results (corresponding to parental division of CD34- KSL cells in media containing SCF + IL-3 13) are shown in grey (mean, 95 % CI) [222]. Results for the simulated cells (transfer pool T) are shown in blue (mean of 50,000 simulation runs). Only cells with complete lineage potential (contribution to all four lineages) are shown. On the x-axis, the lineage contribution of the first daughter is given on top, of the second daughter below (1-neutrophil, 2-megakaryocyte, 3-erythroblast, 4-macrophage).

was evaluated by following the fate of its two daughter cells.

The simulation setup is closely related to the above stated case for the lineage contribution of single cells. In order to account for the separation after the parental division, randomly chosen cells from the respective pools T and S of the *source assay* are transferred in the intermediate *division assay*. The *division assay* is represented by an empty model system mimicking the culture conditions for the division of the parent cell. For simulation efficacy, all transferred cells are under the governance of signaling context Ω. The cell cycle position, the affinity a, and the lineage propensities $\mathbf{x}$ are preserved during the transfer from the *source assay* to the *division assay*. After division, both daughter cells are finally transferred into two separate empty model systems in which the development of the progeny is observed for 240 h (*lineage assay*). Finally, the number and the lineage of cells produced in each lineage assay are evaluated and compared to their sibling.

As in the case of the single cell differentiation assays, Ogawa and coworkers [27] distinguished six different lineages within their series of experiments. For the spleen-derived mouse cells, the authors observed that the majority (73%) of the paired daughter cells contributed to just one lineage, identical for both daughters,

suggesting that the parental cell had already been committed to one particular lineage (*identical single lineage contribution*). In addition, a number of paired daughter cells were observed which contributed to more than one lineage. In 10% of the experiments, both daughter cells contributed to the same combination of lineages (*identical multiple lineages*), whereas in 17 % the daughter cells contributed to different combinations of lineages (*non-identical multiple lineages*). Under the outlined assumptions, the experimental results can be reproduced using exactly the same transfer pool S as it has been used for the single cell differentiation experiments. The comparison of the data is shown in Figure 5.22(A). In close correspondence to additional findings reported in [27], it is observed that for one pair of cells one daughter cell might develop into up to five lineages while the other daughter cell is restricted to just one or two. Furthermore, some simulations generated daughters that contribute to the same overall combination of lineages with considerably different proportions of the individual cell types among their progeny.

For the case of the more sophisticated stem cell sorting procedure used by Takano et al. [222] the narrower transfer pool T is applied. Changing no other parameters, the model reproduces the general results of the experiments: Among the initial parental cells with complete lineage contribution (all four lineages occur in at least on of the daughter colonies) paired daughter cells with identical lineage development dominate over pairs with asymmetric development (Figure 5.22(B)). In the simulation results additional minor contributions (0.1-8.0%) to other combinations of lineages are observed (data not shown). These are not described experimentally, which is most likely due to the limited number of observations.

Parameters for the simulations on paired daughter cells are provided in Appendix D.5.

5.4.3. Lineage specification in differentiating cell cultures

The ability of the simulation model to account for varying fractions of committed cells is verified in the context of a well characterized cell line, namely the FDCP-mix cell line. FDCP-mix cells are derived from murine, multipotent hematopoietic progenitors and retain the capacity to self-renew in the presence of high concentrations of IL-3 [39, 97]. When transferred to low concentrations of IL-3 combined with other hematopoietic growth factors, or injected into experimental animals, FDCP-mix cells show an apparently normal progression of lineage commitment and differentiation. FDCP-mix cells maintained in Iscove's Modified Dulbecco's Medium (IMDM) containing 20% horse serum and 100u/ml Interleukin-3 (IL-3) were washed and transferred at a density of 4 x 104 cells/ml to IMDM containing 20% fetal calf serum and either myeloid (M) or erythroid (E) growth factors as previously described [97]. The combination of growth factors support differentiation either into a mixture of granulocytes and macrophages (M) or into a

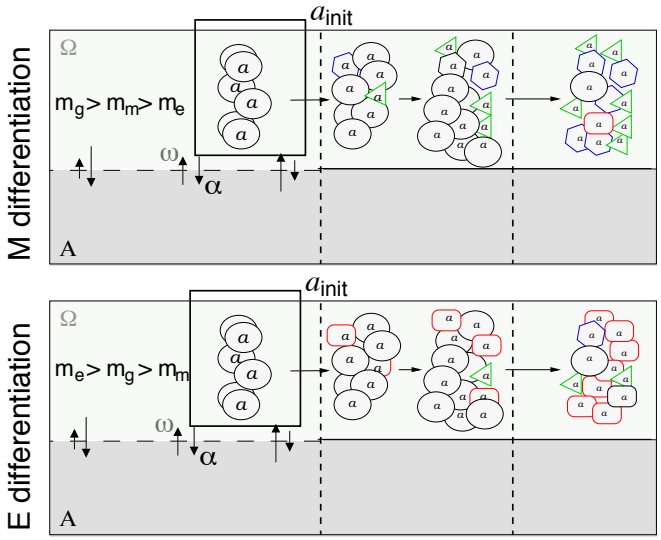

Figure 5.23.: Simulation strategies for differentiating cell cultures.
For the simulation 250 cells are initialized within a well defined initial affinity a, uniformly distributed in the range $a_{\text{init}} = [0.01, 0.1]$ in signal context Ω, indicated by the boxes marked a_{init}. Simulation parameters are individually adapted for the differentiation in granulocyte/macrophage (M) or erythrocyte (E) stimulating conditions.

predominantly erythroid (E) population. On consecutive days up to day 9, cells were harvested from replicate cultures and cytospun. Following May-Grunwald staining, differential counts were performed blind on 100-200 cells per time point. This way, a temporal pattern of the differentiation process could be obtained. The experimental data was kindly provided by Michael Cross.

In order to reflect the usage of relatively homogenous cells from a cell line in the simulation model, the differentiation assay is initialized with a population of 250 cells with a well defined initial affinity, uniformly distributed in the range $a_{\text{init}} = [0.01, 0.1]$, compare Figure 5.23. The fraction of undifferentiated and committed cells is evaluated hourly for a period of 9 days. A balanced expression of the $N = 3$ lineage propensities $x_{1...3}(t = 0) = 1/3$ is assigned to the cells, such that they are initially identical for the development in each of the three experimentally observed cell types: granulocytes, macrophages and erythrocytes. Both, the *instructive* and the *selective* scenario of lineage specification have been applied to analyze the observed development of the cultures under granulocyte/macrophage (M) or erythrocyte (E) stimulating conditions.

To simulate the differentiation of FDCP-mix populations in the presence of growth factors in the *instructive* scenario, the corresponding rewards $m_i = n_i$ of the dissipative control regime are adapted in favor of the relevant fates. The

simulation results are compared to the experimental data in Figure 5.24(A), (B) for both the differentiation in M- and E- media. The lineage specific values for the rewards $m_i = n_i$ in the dissipative regime are the only difference assumed for the different media conditions. During erythroid development, erythrocytes mature from erythroblasts. Since it is possible to distinguish between these cell types morphologically, erythroid cells are subdivided for the phenotypic mapping (cf. Section 4.3.3), such that the committed cell stage now comprises early committed cells (erythroblasts) and mature cells (erythrocytes). Figure 5.24 confirms that the simulation model in the *instructive* scenario is able to quantitatively account for the temporal development of the proportions of observed cell types in both, M- and E-media. Parameter settings for the simulations are provided in Appendix D.6.

Alternatively, the experimental data can be described in the context of *selective* lineage specification, demonstrated in Figures 5.24(C), (D). As discussed in Section 5.2 the lineage specific rewards in the dissipative regime are identical for all three observed lineages ($m_1 = m_2 = m_3$) such that they are all promoted with equal likelihood. However, the simulations for the different culture conditions (M and E) differ in the choice of the lineage specific cell death intensities ϕ_i. Whereas under M conditions the erythroid cells undergo cell death, the situation is different under E conditions in which the selection process leads to the preferential removal of cells committed towards the granulocyte and macrophage lineages.

In both, the *instructive* and the *selective* scenario of lineage specification, the regulating parameters have been adapted to meet the particular experimental results. However, the agreement of simulation and experiment can be regarded as a proof of principle that the proposed model is able to adequately account for differentiation kinetics on the population level. The question remains whether the model could help to distinguish between instructive and selective mechanisms governing the process of lineage specification under particular conditions. As shown above, this is hardly possible on the population level. However, it will be demonstrated in the next chapters that the developmental process on the single cell level are more adequate to address this question.

5.5. Summary and additional remarks

The extensive analysis of the system dynamic and the application to a set of relevant experimental results illustrates that the model successfully accounts for the critical phenomena of lineage specification stated in Section 4.1. In particular the model is suited to account for the central aspects of restriction of lineage potential and the generation of diversity (criteria C1 and C2). This is substantiated by the application of the model to results on the differentiation of single cells and paired progenitors in Sections 5.4.1 and 5.4.2. As the model of lineage specification is

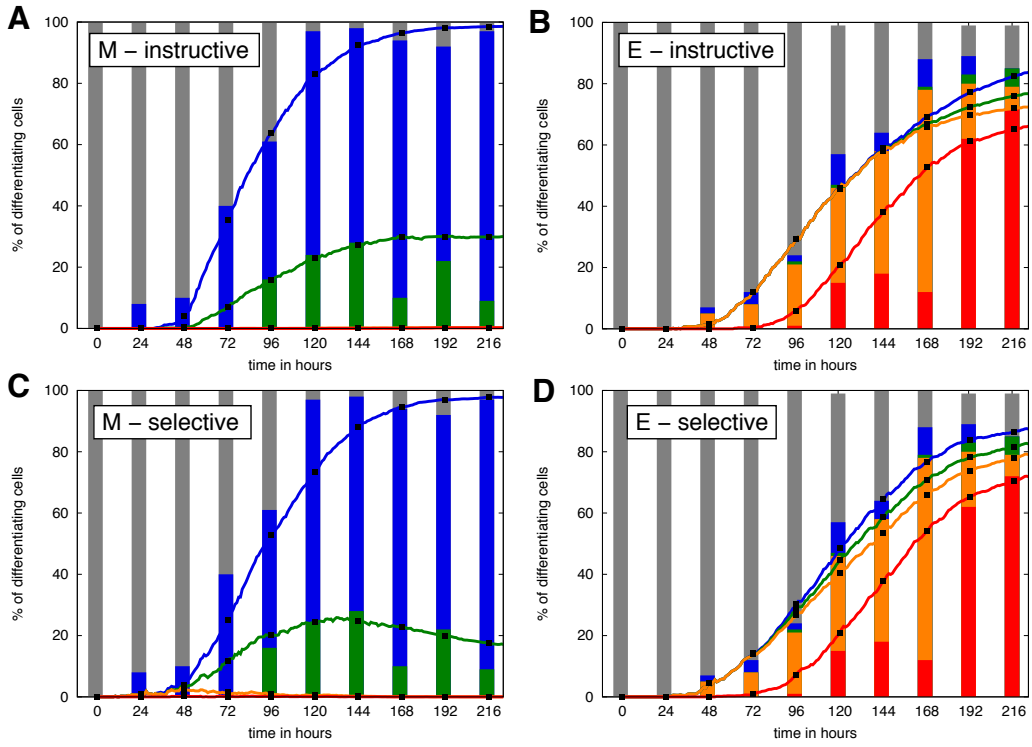

Figure 5.24.: Lineage contribution of FDCP-mix cells.
Experimental results for differentiation in M-medium **(A,C)** and E-medium **(B,D)** are illustrated by the bars (assessed by morphology counts on a daily basis) as a function of time. The simulation results are indicated by the corresponding lines (blue - granulocytes; green - macrophages; orange - committed erythroblasts; red - mature erythrocytes; grey - undifferentiated progenitors). (A,B) correspond to the *instructive* mode of lineage specification, (C,D) to the *selective* mode.

constructed as an "accelerated drift" towards a random lineage rather than an immediate decision the aspects of temporal extension and reversibility are naturally integrated as well (criterion C3). The applicability of an *instructive* and a *selective* mechanism of lineage specification are strikingly demonstrated using the time course data on the differentiation of the FDCP-mix cell line under different culture conditions (criterion C4). However, the close concordance of this two fundamentally different mechanisms to explain the same data set hints towards a structural deficiency of most population based approaches. It is precisely the averaging over a multitude of cells that erases minor differences on the single cell level. Therefore a view is advocated analysing lineage specification not only as a population phenomena but on the basis of individual cellular developments. Using the proposed model the necessary methodology is presented in the subsequent chapters.

It appears as the logical continuation of the outlined research to couple the regulation of lineage specification, this means the regulation of the actual lineage contributions, to the presence and concentration of progenitor and mature cells. It has been shown experimentally as well as theoretically that there exist multiple feedback loops (mostly facilitated by cytokine concentrations) between more mature cell stages and the stem cell compartment [157, 224]. However, the current model lacks a sufficient description of the development of later cell stages. Therefore, the introduction of possible feedbacks would be rather speculative and can not go beyond a general statement. An extension and adaption of the model approach is scheduled to include aspects of cell maturation.

Similar restrictions apply to the study of lineage bias phenomena. It has been shown in a number of experiments that hematopoietic stem cells can obviously inherit a certain preference in their lineage contribution [44, 45]. Addressing this data in the concept of the presented model is subject to another project which is currently underway.

The main results within Section 5.4 have been published in [114]. The model simulations for this publication are based on a linear reward function in the regressive control regime instead of the sigmoid reward function used throughout this thesis (compare Section 4.3.2). However, the results are basically identical as this is a minor change under the applied conditions.

6. Methods II: Characteristics of single cell development

Classical cell culture approaches address cellular proliferation and differentiation potential on the population level. However, the application of time-lapse video monitoring combined with appropriate analysis methods will soon allow to map these observables on the level of single cells. The resulting data structure represents the developmental genealogy of each individual cell, in which all the population measures are encoded. The following two chapters of this thesis are devoted to the statistical analysis and comparison of the resulting data type in a defined mathematical framework.

6.1. Records of single cell development

The stem cell model presented in this thesis is based on the simulation of individual cells. Although the same set of rules apply to all these cells, each and every cell can obey an individual position in the space of their defining variables, i.e. cell cycle position c, attachment affinity a, signaling context $\mathcal{A}$ or Ω, and lineage propensities $\mathbf{x}$. As outlined in Section 3.4 this individual position influences the tendency for certain future developments. The cell-based character of the modeling approach allows studying cell fate decisions on the level of single cells and their progeny and to record all relevant changes in their defining variables. For the example of the cell differentiation assays for the FDCP-mix cell cultures studied in Section 5.4.3, two such records are shown for randomly chosen cells in Figure 6.1. These tree-like structures are more generally referred to as *cellular genealogies*.

As discussed in Section 2.4 a comparable type of data can be obtained by analyzing time lapse videos taken from appropriate cell cultures. By keeping track of the position and the divisional history of each initial cell one obtains a number of unique examples of developmental sequences as they occur under the particular assay conditions. The suitability of this approach has been shown recently in a number of relevant publications [126, 128, 130, 225, 226].

The nature of the data type, i.e. *in situ* records of single cell development, is closely similar between experimental data and the proposed model framework. On one hand the experimental data will contribute to a refinement of the simulation model (e.g. with respect to the distribution of cell cycle times) but on the other hand the model analysis allows a backward interpretation of the data in the light of the model. The identification of similarities between experimental and simu-

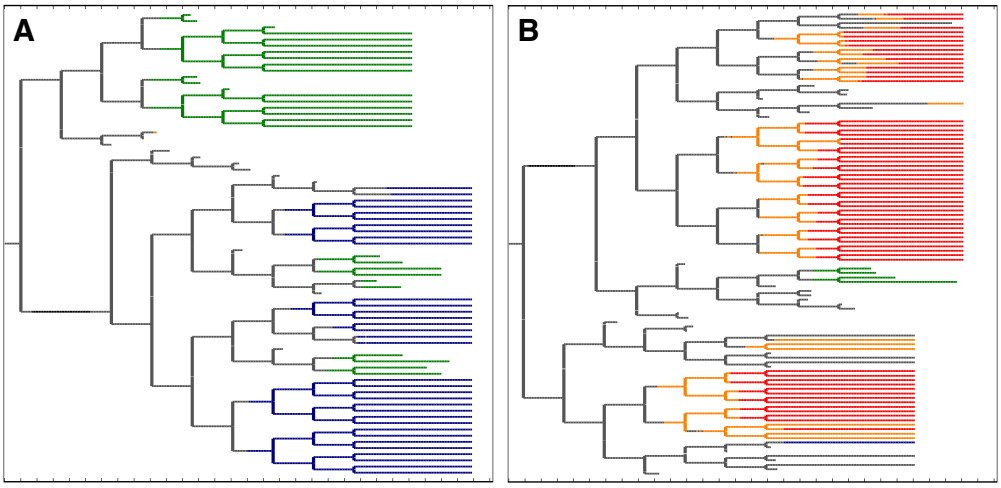

Figure 6.1.: Examples of simulated cellular genealogies.
The two examples shown in **(A)** and **(B)** correspond to single cell records from the simulated population experiments shown in Figures 5.24(C), (D), respectively. The vertical lines represent division events and the horizontal lines represent cells. The color coding corresponds to Figure 5.24, too. Further details of cellular genealogies are described in the main text below.

lated data could reveal more insights about generic principles that govern cellular development. However, due to technical limitations the availability of experimental data is still limited, although it is expected to gain more importance in the next years.

In cell culture experiments the overall expansions and the differentiation potential can be easily compared between different cell culture conditions. However a comparison of the numerous corresponding cellular genealogies is more challenging. In particular, the complexity of the particular type of "tracking data" requires a rigorous characterization and a set of well suited measures for statistical evaluation. Moreover, a large set of such cellular genealogies can reveal typical patterns of cellular development and clonal heterogeneity which are hardly visible by analyzing single realizations or observing the population average. Based on this motivation the following topics are addressed:

- Establishment of a formal characterization of cellular genealogies for the analysis of the divisional history (i.e. topology of the tree-like structure) and for the storage of cell specific and time dependent parameters such as size, morphology, fluorescence activity etc.

- Development of a set of measures for the comparison of different sets of cellular genealogies with a focus on:
  - measures of clonal expansion
  - measures of asymmetry within cellular genealogies
  - measures of correlation between characteristic events
- Comparison of methods for representing cell fate decisions within cellular genealogies.

The introduction of the terminology and the proposed measures within this chapter is accompanied by two examples of their application in the context of the simulation model in the next chapter.

6.2. Formal characteristics of cellular genealogies

Cellular genealogies are derived from the tracking of a single, specified cell object (generally referred to as *root cell*) and its entire clonal offspring. Technically, a cellular genealogy is an unordered tree graph $\mathcal{G} = (\mathcal{C}, \mathcal{D})$ composed of a set of edges $\mathcal{C} = \{c_i,\, i = 0 \ldots n\}$ representing cells and a set of branching points $\mathcal{D} = \{d_j,\, j = 1 \ldots m\}$ representing division events. Unordered trees are characterized as trees in which the parent-daughter relationship is significant, but the order among the two daughter cells is not relevant.

Each genealogy is uniquely identified by its root cell c_0 which is the cell that had been chosen as the initial cell of the tracking process. The root cell and all its descendents are ordered into subsets $\mathcal{C}_g$ according to their generation g, starting with the root cell $c_0 \in \mathcal{C}_0$ and followed by the daughter cells in the first to the gth generation. Furthermore, cells are characterized by their future development, i.e. to each cell c_i belongs either a subsequent division event d_j, giving rise to two daughter cells ($c_i \in \mathcal{C}^{\mathrm{div}}$, with $\mathcal{C}^{\mathrm{div}}$ representing the subset of all cells which undergo division), or the cell's existence terminates without a further division either by cell death ($c_i \in \mathcal{C}^{\mathrm{death}}$, with $\mathcal{C}^{\mathrm{death}}$ representing the subset of all cells which die within the observation period) or by termination of the tracking process ($c_i \in \mathcal{C}^{\mathrm{term}}$, with $\mathcal{C}^{\mathrm{term}}$ representing the subset of all cells with censored observation, i.e. no information about future cell fate is available). Final cells are termed *leaf cells*, i.e. $\mathcal{C}^{\mathrm{leaf}} = \mathcal{C}^{\mathrm{death}} \cup \mathcal{C}^{\mathrm{term}}$. The relation r_{pq} between any two cells c_p and c_q is defined as a topological distance which measures the number of divisions between these cells. Daughter cells that share the same parental cell are termed *siblings*. A schematic representation of a cellular genealogy and an illustration of the distance measure are provided in Figure 6.2.

The temporal dimension of the tracking process is usually encoded in the length of the edges; however this is an associate information rather than a genuine topo-

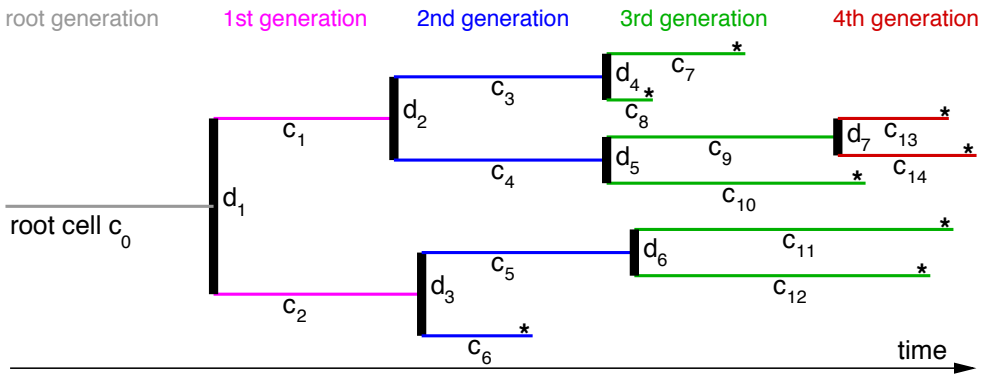

Figure 6.2.: Schematic sketch of a cellular genealogy.
Within the given five generation genealogy the thin horizontal lines represent the cells c_i whereas the divisions d_j are marked by the thick vertical bars. Colors correspond to the cell's generation g. The horizontal dimension is time t with the founding root cell c_0 indicated on the left side (grey, generation $g = 0$). Thus, the length of the horizontal lines represents the duration of the cell's existence and is a measure of the cell cycle time τ_c. Final cells on the right side are called leaf cells (marked with an asterisk). The degree of relation r_{pq} between any two cells c_p and c_q is given by the number of divisions between them. For example, cells c_6 and c_8 have a degree of relation $r_{6,8} = 4$ (separated by the divisions d_3, d_1, d_2, and d_4). Using the same measure of relation, the branch length from the root c_0 to the leaf cells is determined. For the particular example the longest branch is $r_{0,14} = 4$ and the shortest branch is $r_{0,6} = 2$.

logical parameter. Similarly, any additional and potentially time-dependent information that has been recorded during the tracking process can be attributed to the corresponding edges c_i, such as the spatial position, the size of the cells, the expression of certain lineage specific marker genes, or the fluorescence activity of particular cell labels. Specifically, in the case that data on the lineage commitment is available, a fate information $\chi_i(t)$ is assigned to the cell c_i. Different methods for this assignment and detailed examples are presented in Section 6.4.

6.3. Topological measures for cellular genealogies

The quantitative assessment of cellular genealogies is the basis for their comparative analysis. Restricting the view to the parent-daughter relationship one obtains an unordered tree structure that can be characterized by suitable topological measures. This topological structure is the defining feature of all cellular genealogies and is independent of any experiment-specific, additional information (e.g. spatial information, fluorescence measurements) assigned to the tracked cell objects.

One might recall that the analysis of tree like structures has a long tradition in phylogenetics and evolutionary biology (see the historical overview in [227]). Comparing different phylogenetic trees, the influence of external pressure on the evolutionary development is characterized and linked to associated patterns in the tree shape. Although some of the proposed measures within this thesis pick up the idea of shape measures in phylogenetics the general approach in the analysis of cellular genealogies starts from a different point: Whereas in statistical phylogenetics a certain tree structure represents are rather unique set of events typical for a certain species, the analysis of cellular genealogies is based on the comparison of many heterogeneous, albeit similar pedigrees derived under identical culture conditions. Also the interpretation of the typical events like cell death/extinction and division/branching is different for cellular genealogies as for phylogenetic trees, changing the focus to other relevant questions.

The following list of suitable topological measures for cellular genealogies has been derived after extensive analysis of a multitude of potential measures using simulated cellular genealogies. The presented measures focus on a distinction between cellular genealogies that have been derived under different experimental conditions with respect to expansion, symmetry and the occurrence of cell death events. Furthermore, these measures proofed useful to quantify the heterogeneity occurring within a set of genealogies that has been derived under identical conditions. The list is by no means complete and may require extension for different experimental questions.

- **Total number of leaves L and the total number of divisions D.** The total number of leaves L is a suitable measure for the clonal expansion of a particular root cell. The index L counts all cells c_i of a certain genealogy that do not terminate with a further division. The number of divisions D occurring in the same genealogy is equally well suited for estimation of cellular expansion since $D = L - 1$. The example of a cellular genealogy shown in Figure 6.3 has $L = 11$ leaves (marked by the asterisks) and $D = 10$ divisions (indicated by the thick vertical bars). Population averages of these values are closely related to the overall expansion of the cell culture.

  Beyond the average values, the width of the distribution of the number of leaves L (or divisions D, respectively) originating from different cells under the same culture conditions is an indicator of population inherent heterogeneity in the clonal expansion potential that cannot be determined on the population level alone.

  Formally, the total number of leaves L is given as:

$$L = \sum_{c_i \in \mathcal{C}} I_D, \text{ with } I_D = \begin{cases} 1 & \text{for } c_i \in \mathcal{C}^{\text{term}} \\ 0 & \text{else} \end{cases} \tag{6.1}$$

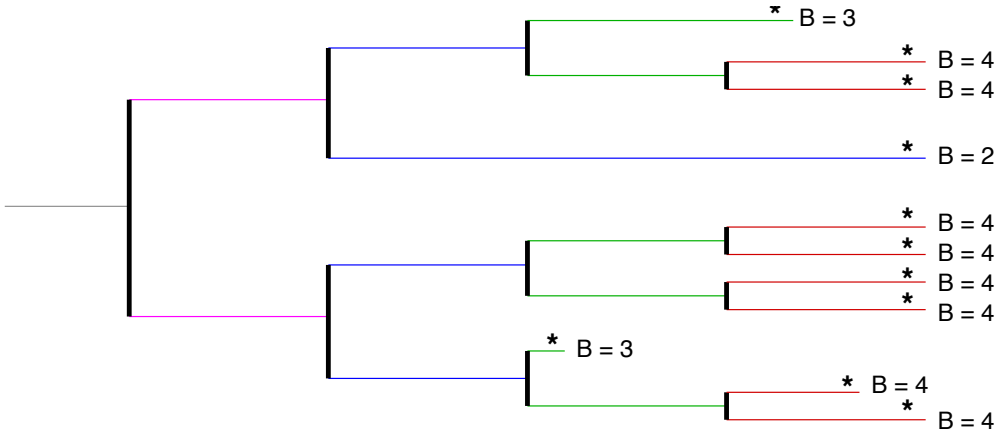

Figure 6.3.: Leaves and branch length.
Example of a cellular genealogy with $L = 11$ leaves (marked by the asterisks) and $D = 10$ divisions (indicated by the black vertical bars). For each leaf cell the branch length B is provided. Color coding of the cells corresponds to the generation g (root - grey, $g = 1$ - magenta, $g = 2$ - blue, $g = 3$ - green, $g = 4$ - dark red).

- **Branch lengths B.** The branch length $B = B_k$ measures the number of divisions between the root cell c_0 and a specific leaf cell c_k. The complete set of branch lengths for all leaf cells of a given genealogy is a measure of the proliferative activity of the root cell, but, it also accounts for the heterogeneity within a single expanding clone. Both these aspects are briefly discussed.

  Due to the exponential nature of cellular expansion, the average branch length $\bar{B}$ within a particular genealogy is dominated by the maximal branch lengths $\max(B)$. To circumvent this inherent bias, a characteristic branch length of a genealogy B^{char} is proposed for which the different branch lengths B_k are weighted by the generation g in which the leaf cell occurs. Intuitively speaking, B^{char} is the expectation value of the branch length by randomly following the genealogy from the root cell c_0 to the leaves. Such a normalization process ensures that longer and more ramified branches are weighted less compared to shorter branches.

  The distribution of branch lengths B_k within a particular genealogy characterizes the heterogeneity within the progeny of a single expanding (root) cell. However, these distributions are always dominated by the longer branches due to the exponentially increasing number of leaf cells. Therefore, it can be argued that the relation between the extreme values $\min(B)$ and $\max(B)$ is more instructive. Defining the range of branch lengths B^{range} as the dif-

ference between the minimal and the maximal branch lengths provides a simple measure to quantify this heterogeneity.

Formally, the branch length B_k is defined as the topological distance $r_{0,k}$ between the root cell c_0 and a leaf cell $c_k \in \mathcal{C}^{\text{term}}$. The characteristic branch length of a genealogy B^{char} is calculated as

$$B^{\text{char}} = \sum_{c_k \in \mathcal{C}^{\text{term}}} (B_k / 2^{g_k}) \tag{6.2}$$

in which g_k refers to the generation of leaf cell c_k. The range of branch lengths B^{range} for a certain genealogy is given as $B^{\text{range}} = \max_k(B_k) - \min_k(B_k)$.

For the example of the cellular genealogy depicted in Figure 6.3 the branch lengths B are provided for each individual leaf cell. For this genealogy, the characteristic branch length evaluates to $B^{\text{char}} = 1 \cdot \frac{2}{2^2} + 2 \cdot \frac{3}{2^3} + 8 \cdot \frac{4}{2^4} = 3.25$ and the range of branch lengths to $B^{\text{range}} = 4 - 2 = 2$.

- **Symmetry index (weighted Colless' index C^w).** Tree shape measures with a focus on symmetry have a long tradition in the analysis of phylogenetic trees [227, 228, 229]. These measures are commonly used to detect imbalances that testify the regulation of diversity in ecological communities. Applied to the situation of cellular genealogies these measures can provide an understanding of the balance between self-renewal and differentiation, as well as on the action of cell death processes.

  A particular useful measure is the Colless' index of imbalance C [230]. This index compares the number of leaves emerging from the two daughter cells $c_{\text{daughter;1}}$ and $c_{\text{daughter;2}}$ resulting from the division of the parent cell c_i. Colless' index C sums the difference in the number of leaves subtended by the two daughter cells for all divisions within the genealogy and normalizes by dividing with the largest possible score. Colless' index increases from $C = 0$ for perfectly symmetric genealogies to $C = 1$ for completely asymmetric genealogies. However, the classical Colless' index puts the same weight on asymmetries that occur late in development compared to earlier events. This is contrary to the common biological perspective on the balance between stem cell self-renewal and differentiation which assumes that asymmetries are most pronounced in the early divisions. Especially in the case of large, exponentially expanding genealogies, such early events are underestimated by the classical Colless' index compared to a vast amount of expansion events in latter stages of development. Therefore, a weighted Colless' index C^w is proposed explicitly accounting for the exponential expansion within cellular

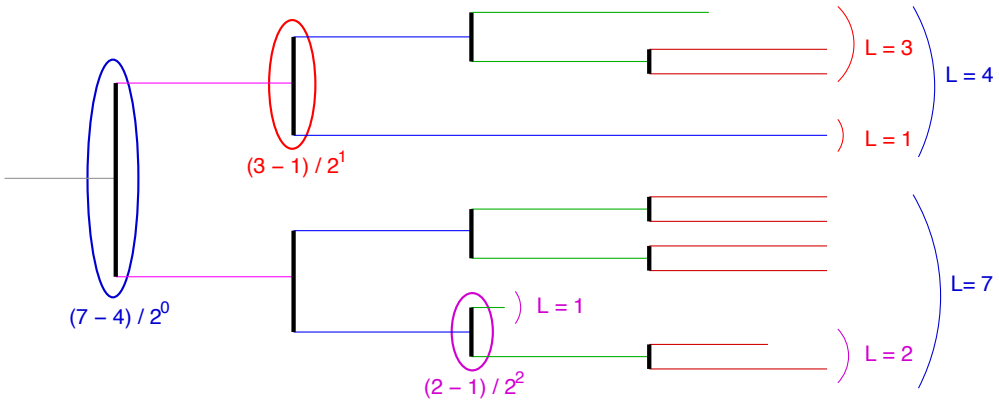

Figure 6.4.: Colless' index.
For three encircled divisions the contributions to the weighted Colless' index C^w are outlined below each event. The color-coded pairs of parenthesis on the right side summarize the number of leaves subtending from each of the relevant daughter cells. The exponent of the denominator corresponds to the generation g of the mother cell. The coloring scheme of the cells indicates their generation g (compare Figure 6.3).

genealogies. In contrast to the classical Colless' index C, the weighted Colless' index C^w sums over the differences in the number of leaves emerging from two daughter cells which are weighted according to the generation in which the asymmetry occurs.

Formally, the classical Colless' index of imbalance is given as

$$C = \frac{2}{(L-1)(L-2)} \sum_{c_i \in \mathcal{C}^{\text{div}}} |L_{i;1} - L_{i;2}|. \tag{6.3}$$

$L_{i;1}$, $L_{i;2}$ refer to the number of leaves subtended by the two daughter cells of cell c_i. In contrast, the weighted Colless' index C^w is given as

$$C^w = N_C^{-1} \sum_{c_i \in \mathcal{C}^{\text{div}}} (1/2^{g_i}) |L_{i;1} - L_{i;2}| \tag{6.4}$$

The generation g_i refers to the generation of the parental cell c_i. C^w is normalized with a constant N_C which corresponds to the maximal possible value of the Colless' index for a completely asymmetric tree with L leaves. N_C is defined as

$$N_C = \sum_{j=0\ldots(L-3)} \frac{(L-2-j)}{2^j} \tag{6.5}$$

For the example of the cellular genealogy in Figure 6.4 contributions to the weighted Colless' index C^w are outlined for the encircled division events. The denominator results from the weighting with the generation g of the mother cells.

- **Cell death index A.** Programmed cell death (also referred to as apoptosis) potentially plays an important role in regulation of hematopoiesis *in vivo* (compare Section 2.3.2) and is also regularly observed in cell cultures. The cell death index A measures the observed frequency of cell death events and is, therefore, an estimate of the probability of cell death occurrence. To account for systematic effects related to cellular development, it seems appropriate to consider the cell death index A as a function of the current cell state and/or the generation g within the genealogy. Therefore, the cell death index A^g is calculated as the ratio of the number of cell death events observed for cells in generation g and the number of all cells existing in the same generation.

  In contrast to the cell death index A^g itself, a generalization to pairs of sibling cells allows to identify potential correlations of cell death events and, therefore, to reveal particular asymmetries in cell fates. The idea behind this approach is that in case of statistically independent events, the probability of observing a particular combination of events in two siblings (i.e. cell death in none, one or both siblings) equals the product of the probabilities of the corresponding events for individual cells. Thus, if cell death events would occur independently of each other, the latter probabilities could be estimated by $(1-A^g)^2$, $2A^g(1-A^g)$, $(A^g)^2$, respectively. Using the differences in the observed and the (under the independence assumption) expected frequencies of these pair-wise events, it is possible to calculate the so called mutual information (MI) of all sibling pairs within a particular generation. The MI, which always has values between 0 and 1, is a measure of the information about one of the two events that is provided by the other one. In the particular case, $MI = 0$ would imply that one cannot obtain any information about the cell death occurrence of one sibling cell from knowing the fate of the corresponding daughter cell, as expected under the applied model assumption of completely random cell death.

  Formally, the cell death index A^g is an estimator of the probability for a cell death event occurring in generation g of a certain genealogy. It is calculated as

$$A^g = \frac{\sum_{c_i \in \mathcal{C}} I_D}{\sum_{c_i \in \mathcal{C}} J_D} \tag{6.6}$$

  in which the indicator function $I_D = \begin{cases} 1 & \text{for } c_i \in \{\mathcal{C}^{\text{death}} \cap \mathcal{C}^g\} \\ 0 & \text{else} \end{cases}$ is used to

count the number of cell death events in generation g and the indicator function $J_D = \begin{cases} 1 & \text{for } c_i \in \mathcal{C}^g \\ 0 & \text{else} \end{cases}$ to determine the total number of cells that exist in the same generation g.

The mutual information of two (discrete) random variables X and Y is defined as

$$MI(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{6.7}$$

in which $p(x,y)$ is the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are the marginal distributions of X and Y, respectively. I.e., the MI is the expected log-likelihood difference between the bivariate model and the product of the marginal models. In the particular case of cell death events, one assumes identical probability distributions for both sibling cells. Therefore, the expected probabilities for the three possible events (i.e. none (p_0), one (p_1) or two (p_2) cell death event per sibling pair) under the hypothesis of statistical independence of the two siblings can be estimated by $(1-A^g)^2$, $2A^g(1-A^g)$, $(A^g)^2$, respectively. Estimating the bivariate probabilities by the observed relative frequencies (f_i, $i = 0, 1, 2$) of the aforementioned events (p_i, $i = 0, 1, 2$), leads to the estimated mutual information per generation g:

$$MI^g = f_0 \log_2 \left(\frac{f_0}{(1-A^g)^2} \right) + f_1 \log_2 \left(\frac{f_1}{2A^g(1-A^g)} \right) + f_2 \log_2 \left(\frac{f_2}{(A^g)^2} \right) \tag{6.8}$$

For illustration of the MI measure, two genealogies are shown in Figure 6.5. Although these genealogies have identical values for the number of leaves L, the characteristic branch length B^{char} and the total number of death cells they differ significantly with respect to correlation of the cell death events. Whereas in Figure 6.5(A) the cell death events are rather isolated, they always appear among sibling cells in Figure 6.5(B). This leads to increased values of the mutual information measure MI in the latter scenario.

It should also be noted that this approach can be generalized to other events characterizing the fate of sibling cells. A related but less analytical approach to correlate fluorescence expression between closely related cells, indicating synchronized epigenetic remodeling in embryonic stem cells, has been published recently [130].

- **Minimal distance between characteristic events R.** Cellular genealogies retain information about the relatedness of certain characteristic cellular events like the occurrence of cell death, changes in the cells morphology, or the expression of cell fate characteristic markers. Beyond the mutual information MI, the topological distance r_{ij} between such characteristic cellular

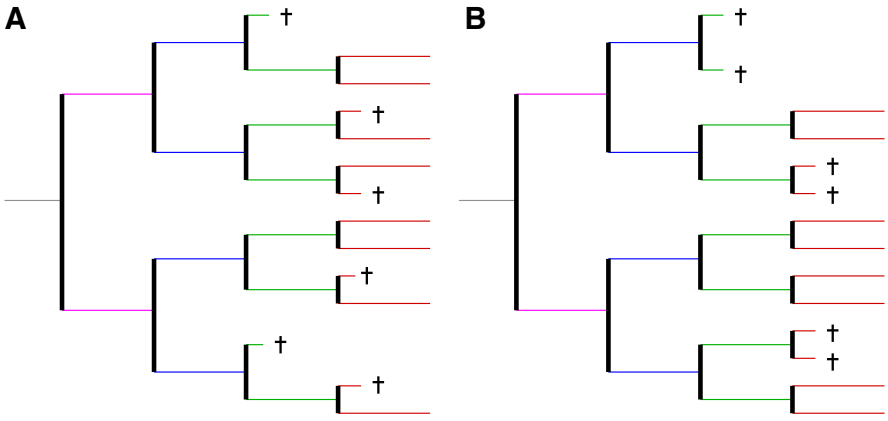

Figure 6.5.: Mutual relation between cell death events.
The genealogies in **(A)** and **(B)** have similar topologically features ($L = 14$, $B^{\text{char}} = 3.75$, $B^{\text{range}} = 1$, $C = 0.05$, $A^3 = 1/4$, $A^4 = 1/3$ for both genealogies). However they differ with respect to occurrence of cell death events. Whereas in (A) the cell death events are rather isolated, they always appear in sibling cells in (B), resulting in differences for the weighted Colless' index ($C^w = 0.068$ in (A); $C^w = 0.136$ in (B)), the mutual information measures MI^g ($MI^3 = 0.036$ and $MI^4 = 0.076$ in (A); $MI^3 = 0.244$ and $MI^4 = 0.276$ in (B)) and the minimal distance R ($R = 3$ in (A); $R = 1$ in (B)). Coloring of the cells indicates their generation g (compare Figure 6.3)

events occurring in cells c_i and c_j has been identified as a suitable measure of their relation. In particular, it is the *minimal* distance (denoted as R_i) between a characteristic event of cell c_i (e.g. cell death) and the closest similar event of another cell c_j that proved useful for the identification of whether the events are rather isolated or appear closely related. If at least two characteristic events occur within a given cellular genealogy, a minimal distance R_i can be calculated for each of them. To provide a unique measure for each cellular genealogy, the average R over these individual minimal distances R_i is calculated separately for each genealogy. Lower minimal distances R indicate a closer relation between the events, possibly due to similar developmental stages of the cells in question, whereas a tendency towards higher minimal distances is more likely caused by general effects independent of the cell state.

Formally, the individual minimal distance R_i for a characteristic event occurring for cell c_i is defined as

$$R_i = \min_{c_j \in \mathcal{C}^{\text{char}}} (r_{ij}) \tag{6.9}$$

in which $\mathcal{C}^{\text{char}}$ refers to the set of cells for which a characteristic event has

been observed and r_{ij} is the topological distance between them. The index R_i (and consequently the average R) is not defined for genealogies with less than two such characteristic events in $\mathcal{C}^{\text{char}}$.

An illustrative example for the detection of closely related cell death events is provided in Figure 6.5. Although the cellular genealogies have the same number of cell death events, the individual distance to the next event is always $R_i > 1$ for the cellular genealogy in Figure 6.5(A) but $R_i = 1$ for all cell death events in Figure 6.5(B).

6.4. Assignment of lineage fates

The limitations in accessing the "commitment state" of a cell *in situ* clearly effects the assignment of a the fate information χ_i to the corresponding cell c_i. Approaching this phenomenon from the model perspective is a good option to appreciate the two fundamentally different views on lineage assignment and their implications for the interpretation of cell division events.

Within the proposed simulation model, lineage specification is represented as a continuous process progressively restricting the number of available developmental options. A simple phenotypic mapping, introduced in Section 4.3.3 allows attributing cells as uncommitted or committed to a particular cell type although a small but continuously decreasing probability for conversion remains. This information about the "commitment state" of a cell is available throughout the whole tracking process for each individual cell. Therefore, it can be represented in the cellular genealogies in a straight forward fashion, which is referred to as the *prospective view*: Applying the outlined mapping procedure, a cell c_i is marked according to its current internal propensity levels $\mathbf{x}(t)$ as undifferentiated $\chi_i(t) = 0$ (for $x_j^*(t) < x_{\text{com}}$) or committed to a certain lineage fate $\chi_i(t) = 1, 2, \ldots, N$ (for $x_j^*(t) > x_{\text{com}}$) with N denoting the number of possible lineages. Figure 6.6(A) shows a typical cellular genealogy in which lineage information is assigned in the *prospective view*.

All divisions d_j within this view are characterized by comparing the lineage specification state of the parent cell (prior to division) to the state of the daughter cells (immediately after division). This results in two classes of division events: *undifferentiated symmetric divisions* if an undifferentiated parent gives rise to two undifferentiated daughters ($\chi_{\text{parent}} = \chi_{\text{daughter;1}} = \chi_{\text{daughter;2}} = 0$) and *symmetric divisions* if a committed parent gives rise two daughters of the same fate ($\chi_{\text{parent}} = \chi_{\text{daughter;1}} = \chi_{\text{daughter;2}} > 0$). Since cell divisions in the underlying model system do not involve any differentiation event and are thus symmetric by definition, asymmetric divisions do not occur in the *prospective view*.

In contrast to the simulation model, lineage assignment is a difficult task in the experimental situation, especially if the cellular genealogy needs to be main-

tained. Using classical time lapse microscopy of a differentiating cell culture, the only, currently available, non-invasive method for this assignment is the identification of cell type specific changes in the cell's morphology. However, changes in morphology are hard to identify and occur rather late compared to changes in the transcriptional activity of cell fate specific genes. Novel techniques that are already developed for hematopoietic stem and progenitor cells [131], allow the targeted placement of genes coding for the expression of fluorescence proteins under the control of particular lineage specific promoters ("reporter genes"). By use of these reporter genes it should be possible to obtain information about the lineage decisions during the tracking process. Currently, there are only preliminary studies using this technique in the context of single cell tracking approaches; however it is the most promising strategy for the prospective assignment of lineage fates in a cellular genealogy.

An already applied technique for the identification of lineage fates in cellular genealogies relies on staining methods. This approach requires the conservation of the final, spatial configuration of the tracking procedure in order to allow a unique mapping into the genealogy. This is only feasible for adherent cell cultures as they are used e.g. for the tracking of neural stem and progenitor cells. However, this assignment of lineage fates refers only to the final configuration and earlier decision events have to be estimated in a *retrospective view*. Given that a lineage fate χ_i is assigned to each leaf cell, the fate of all cells within the genealogy is determined recursively as follows: If both daughter cells of a parental cell belong to the same lineage, then the same lineage is attributed to the parent cell ($\chi_{\text{parent}} = \chi_{\text{daughter;1}} = \chi_{\text{daughter;2}}$). The particular division is characterized as *symmetric*. In contrast, if the daughter cells are of different lineages or one is undifferentiated ($\chi_{\text{daughter;1}} \neq \chi_{\text{daughter;2}}$), then the parent cell is marked as undifferentiated ($\chi_{\text{parent}} = 0$) and the parental division is counted as *asymmetric*. Two undifferentiated daughter cells ($\chi_{\text{daughter;1}} = \chi_{\text{daughter;2}} = 0$) derive from an undifferentiated parent ($\chi_{\text{parent}} = 0$) due to an *undifferentiated symmetric division*. Evaluating the same cellular genealogy as Figure 6.6(A) in the *retrospective view* (i.e. only based on the lineage fate of the leaf cells) a modified version of the genealogy is obtained as shown in Figure 6.6(B). There, the progeny of a parental cell that only gives rise to one cell lineage is always shown in the same color.

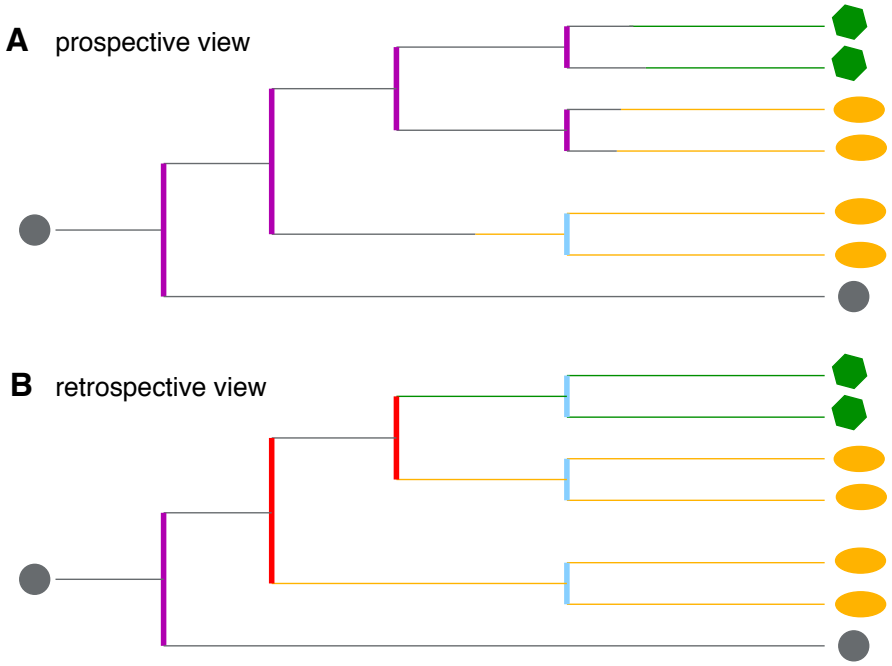

Figure 6.6.: Prospective versus retrospective view for the lineage assignment.
(A). The sketch shows an example genealogy for which the lineage fate is assigned in the *prospective view*. Coloring of the cells indicates the state of commitment as given by the phenotypic mapping, i.e. "the color coding" of a cell might change during the cells existence from uncommitted (grey) to either the "orange" or the "green" lineage depending on the dominating lineage propensity x_i^*. In contrast, for the identical genealogy in **(B)** the lineage fate is assigned recursively based on the lineage fate of the daughter cells (i.e. backwards from the final configuration shown on the right). Colors for the divisions are assigned as follows: undifferentiated symmetric divisions - magenta, symmetric divisions of committed cells - light blue, asymmetric divisions (only in the *retrospective view*) - red.

7. Results II: Quantitative analysis of 'in silico' cellular genealogies

Within this chapter, the statistical measures for cellular genealogies and the different approaches for the lineage assignment are applied to different sets of reference data. Since practical problems with the generation of sufficiently long and qualitatively analyzable time lapse videos of suitable cell cultures as well as difficulties in the automatic identification and tracing of single cells in current image-processing techniques limit the availability of experimentally derived cellular genealogies, the reference data is taken from the single-cell based model introduced in the previous chapters. Based on this model, it can be analyzed how changes in the particular (*in silico*) growth conditions influence the topology of the cellular genealogies and how the imprinted mechanisms of lineage specification can be reextracted from the topology information.

7.1. Comparing cellular genealogies under different growth conditions

7.1.1. Generation of cellular genealogies under different growth conditions

In order to test the different measures for the quantitative characterization of cellular genealogies three different "*in silico* conditions" are studied which are inspired by typical cell growth scenarios.

- First, an empty model system is initialized with one "model stem cell" which undergoes massive expansion. This is referred to as the *growth scenario*.
- Thereafter, the model system establishes a stable pool of self-renewing cells that simultaneously contribute to a pool of differentiating cells. This is referred to as the *homeostatic scenario*.
- Changing the system parameters such that the self-renewal ability of the cells is lost, the whole population of cells undergoes final differentiation and subsequent cell death. This is referred to as the *differentiation scenario* which is inspired by *in vitro* cultures of stem and progenitor cells lacking self-renewal promoting conditions.

Lineage specification is realized such that each of the $N = 3$ possible lineage fates occurs with the same probability (i.e. balanced rewards $m_i = n_i = m$). For the derivation of the cellular genealogies in the *growth scenario* 400 independent model realizations are tracked for 300 hours (one hour corresponds to one time step of the simulation), each initialized with one single stem cell. Only in this situation with an empty model system, the strong initial expansion is observed. In contrast, for the *homeostatic scenario* and for the *differentiation scenario* all cells in the homeostatic stem cell compartment of *one* particular model realization are uniquely marked and subsequently tracked for the next 300 hours. Typically around 400 cells are tracked in this process, similar to the 400 independent realizations in the *growth scenario*. For the *homeostatic* and *differentiation scenario*, there is no statistical difference in choosing individual cells in different realizations or within the same realization as long as the systems are initially in the homeostatic situation (i.e. stable numbers of stem cells in signaling context $\mathcal{A}$ and Ω). A schematic representation of the cell population dynamics for the different scenarios and a typical characteristic cellular genealogy for each scenario is shown in Figure 7.1. Further parameter settings for these simulations are provided in Appendix D.7.

For the exploration of the effects of cell death on the cellular genealogies, a background cell death is applied with constant intensity $\Phi^B = \phi$ for all cells in G1-phase (cf. Section 4.3.4). Generally, such an effect might also occur in other stages of the cell cycle. However, the focus of this analysis is directed towards the (quantitative) characterization of the general impact of cell death events on cellular genealogies rather than on the details of the biological process. The simplifying assumption of restricting cell death events to G1-phase does not qualitatively change the results wherefore the analysis is restricted to this scenario without loss of generality. A discussion of selective cell death as a regulating process in lineage specification (*selective lineage specification*, compare Sections 4.3.4 and 5.2.3) and its effect on the resulting cellular genealogies is provided Section 7.2.

7.1.2. Topological characterization and comparison

The topological measures introduced in Section 6.3 are now applied to the sets of cellular genealogies resulting from the above stated simulation scenarios. It is the central advantage of this comprehensive set of measures to allow for the quantitative comparison of the different cellular genealogies which goes far beyond the visual impression as it is provided in Figures 7.1(B-D).

As some of the measures are not invariant under changes of the observation period their scaling behavior needs to be discussed separately. This is especially necessary in the case, in which genealogies from experiments with different observation periods need to be compared. Already in the model situation, saturation effects (in the *growth scenario*) or exhaustion (in the *differentiation scenario*) are

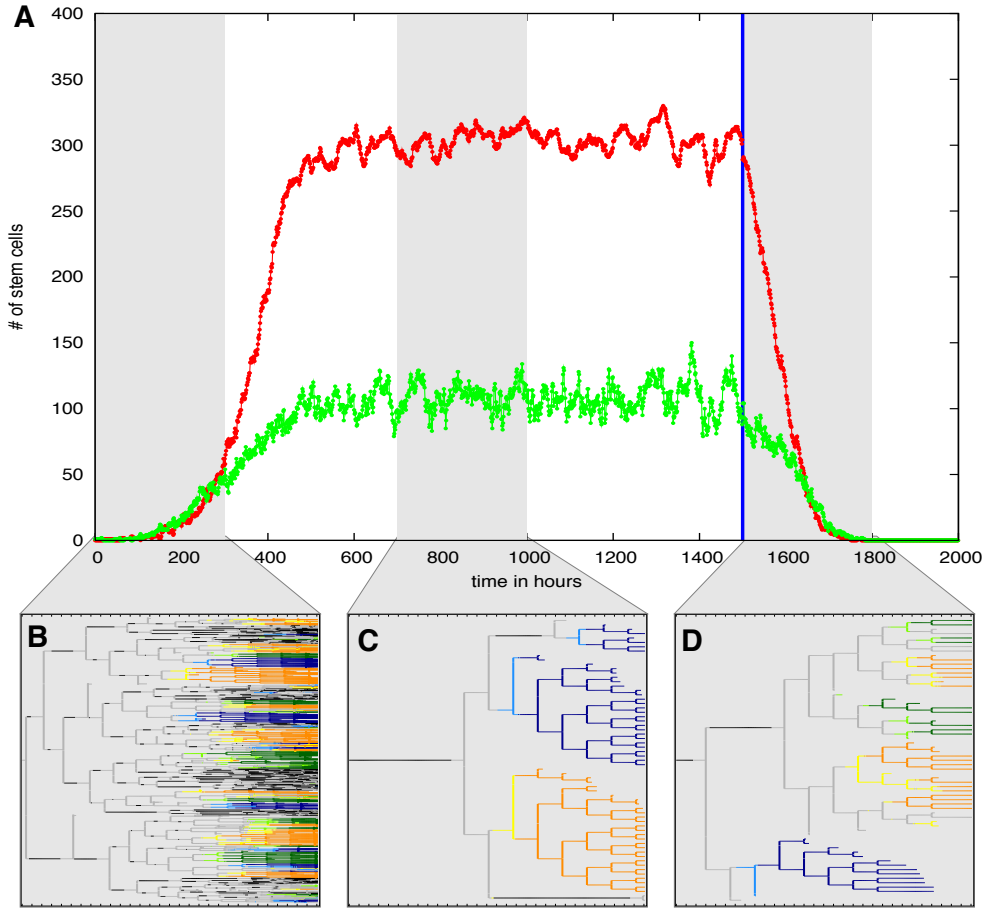

Figure 7.1.: Simulation scenarios and corresponding cellular genealogies. **(A).** Numbers of stem cells in Ω (green) and $\mathcal{A}$ (red). At time t=0, the system is initialized by a single stem cell, that subsequently undergoes expansion. The corresponding *growth scenario* is indicated by the first shaded area (observation period of 300 hours, tracking of 400 individual cellular genealogies in independent realizations). Around t=600 the system reaches a dynamically stabilized equilibrium. For the cellular genealogies of the *homeostatic scenario*, all stem cells present at time point t=700 are uniquely marked and subsequently tracked for 300 hours (second shaded area). Changing differentiation and regeneration parameters at time t=1500 (blue line), the self-renewal ability of the stem cells is lost and they undergo terminal differentiation (*differentiation scenario*). As in the *homeostatic scenario*, cellular genealogies are derived by marking all stem cells present at time point t=1500 prior to the change of parameters and their subsequent tracking for 300 hours (third shaded area). **(B - D).** Characteristic genealogies for each scenario ((B) - *growth scenario*, (C) - *homeostatic scenario*, (D) - *differentiation scenario*). Colors indicate cell cycle status of the undifferentiated cells and commitment to three possible lineages for the differentiating cells: grey - undifferentiated proliferating cell (Ω); black - undifferentiated quiescent cell ($\mathcal{A}$); yellow/orange - early/finally committed cell of the "orange" lineage; light/dark blue - early/finally committed cell of the "blue" lineage; light/dark green - early/finally committed cell of the "green" lineage.

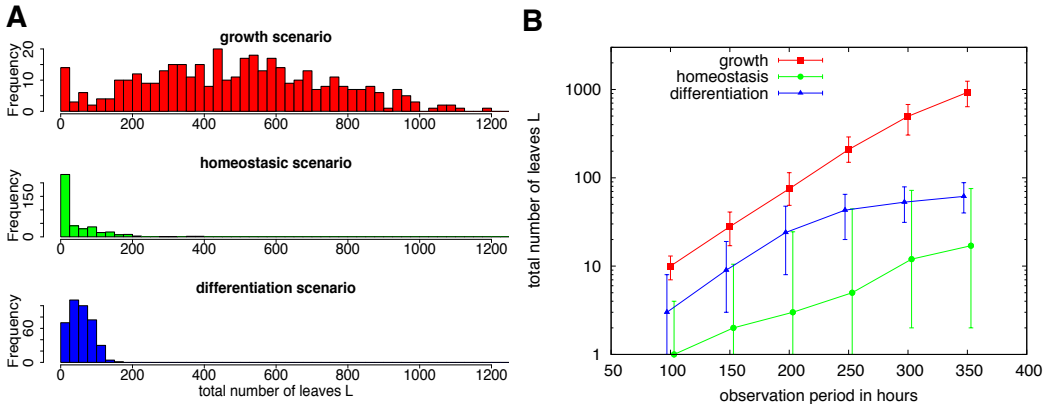

Figure 7.2.: Total number of leaves L.
(A) shows histograms of the distributions of the total number of leaf cells L for the three different scenarios *growth, homeostasis* and *differentiation*. **(B)** provides corresponding scaling behavior, i.e. median of the number of leaves L (indicated by the dots and supplemented by the first and third quartiles) as a function of the observation period (ranging from 100 h to 350 h, shown on x-axis).

not negligible and can lead to a nonlinear divergence from the anticipated behavior. It can be expected that the influence of such effects is even more pronounced in the experimental situation. Therefore, an additional discussion of the particular scaling behavior for each measure is provided in the context of the model.

Total number of leaves L. The total number of leaves L has been introduced as a measure to quantify clonal expansion.

Histograms of the distributions of the total number of leaves L for the three scenarios (*growth, homeostasis* and *differentiation*) are provided in Figure 7.2(A). Increased values of L in the *growth scenario* are plausible since the initial expansion is characterized by high proliferative activity and a shortening of the effective cell cycle time[1], which leads to an increased number of cell divisions during the observation period for the cellular genealogies. In contrast, the *homeostatic* and the *differentiation scenario* show only moderate expansion as the effective cell cycle time returns to normal values.

For the validation of the scaling of the topological measures, further sets of cellular genealogies are obtained which use the same set of root cells but for which the observation time is varied, ranging from 100 hours to 350 hours. In

[1]Due to low cell numbers in both signaling contexts $\mathcal{A}$ and Ω, cells in Ω are quickly attracted to change into $\mathcal{A}$ after division. However, these cells are also quickly reactivated into Ω to undergo the next round of divisions, thus leading to an effective shortening of the average cell cycle times. For further details the reader is referred to [47].

the case of unlimited growth, one would argue the total number of leaves L scales exponentially with the observation period. As shown by the log-lin plot in Figure 7.2(B) this scaling behavior is verified for wide ranges of the observation period. The different slopes indicate different overall expansion rates. A saturation occurs for the *differentiation scenario* which is caused by the limited proliferative activity of the differentiating cells within the model. The variance in the number of leaf cells L is indicated by the error bars (corresponding to the interquartile range) for the different sets of cellular genealogies. As this is also transformed to the log-lin scale the variance can only be compared on a relative scale.

Branch lengths B. Similar to the total number of leaves L the characteristic branch length B^{char} addresses the expansion of a cell clone within a given time interval.

Histograms of the characteristic branch length B^{char} for the cellular genealogies derived under the three different culture scenarios are shown in Figure 7.3(A). It is evident that the characteristic branch length B^{char} is increased in the *growth scenario* as compared to the other scenarios. As illustrated above, this effect is induced by the shortening of the effective cell cycle times as the system has not reached its equilibrium state, thus allowing more divisions within a fixed time period. The dominant peak at very short branch lengths in the *homeostatic scenario* derives from a population of cells mainly residing in signaling context Ω rarely undergoing cell division at all. In comparison, almost all cells in the *differentiation scenario* are activated into cell cycle and undergo regular divisions. Therefore, the observed heterogeneity in the observed characteristic branch length B^{char} is rather limited.

As expected, Figure 7.3(B) supports the notion that the characteristic branch length B^{char} scales linearly as the observation period varies. This scaling behavior results from the direct coupling of the number of subsequent cell divisions to the duration of the observation. Possible saturation effects that are expected for experimental realizations are most likely caused by limitations of space, nutrition and proliferation capacity of the cells.

Both, the total number of leaves L and the characteristic branch length B^{char} are measures for the expansion of cell cultures under different conditions. Although the total number of leaves L scales exponentially with time and the characteristic branch length B^{char} scales linearly with time, the resulting histograms in Figures 7.2(A) and 7.3(A) are qualitatively related. However, on top of this "classical population measure" to quantify the expansion of cell cultures, the individual cellular genealogies give access to the heterogeneity within a cell population. In this sense it is the variance of the total number of leaves L and of the characteristic branch length B^{char} which indicates that individual cells undergo expansion at a very different extent.

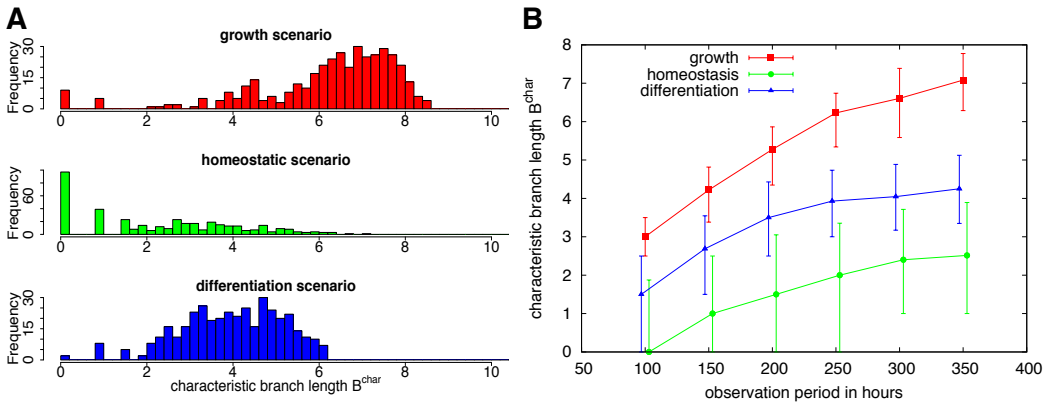

Figure 7.3.: Characteristic branch length B^{char}.
(A) shows histograms of the distributions of the characteristic branch length B^{char} for the three artificial culture scenarios. **(B)** provides median values of the characteristic branch length B^{char} (indicated by the dots and supplemented by first and third quartiles) for different observation periods (ranging from 100 h to 350 h).

The range of branch lengths B^{range} is a possible measure to address this heterogeneity since it quantifies the difference between the shortest and the longest branch within one genealogy. Histograms for the corresponding distributions are shown in Figure 7.4(A). The dominance of high absolute values in the *growth* and in the *differentiation scenario* indicates that uniform expansion in all branches is rarely observed and that the genealogies are characterized by significant differences in the branch lengths B within individual genealogies. This effect is less pronounced in the *homeostatic scenario*. However, the increased occurrence of cellular genealogies with low characteristic branch length $B^{\text{char}} \approx 0$ (i.e. cells with extended periods of cellular quiescence, compare Figure 7.3(A)) manipulates this perspective.

The range of branch lengths B^{range} scales linear with the observation period, see Figure 7.4(B). This scaling results from the linear scaling of the maximal branch length $\max(B)$ whereas the minimal branch length $\min(B)$ reaches a constant value for sufficiently long observation periods due to the action of background cell death. Again, the limitations in the proliferative activity of the differentiating model cells lead to a saturation effect in the *differentiation scenario*.

Symmetry index (weighted Colless' index C^w). Colless' index C and the weighted Colless' index C^w address the question of how proliferation and quiescence are balanced on the level of individual cells. However, application of the classical Colless' index C requires careful interpretation since all asymmetries are

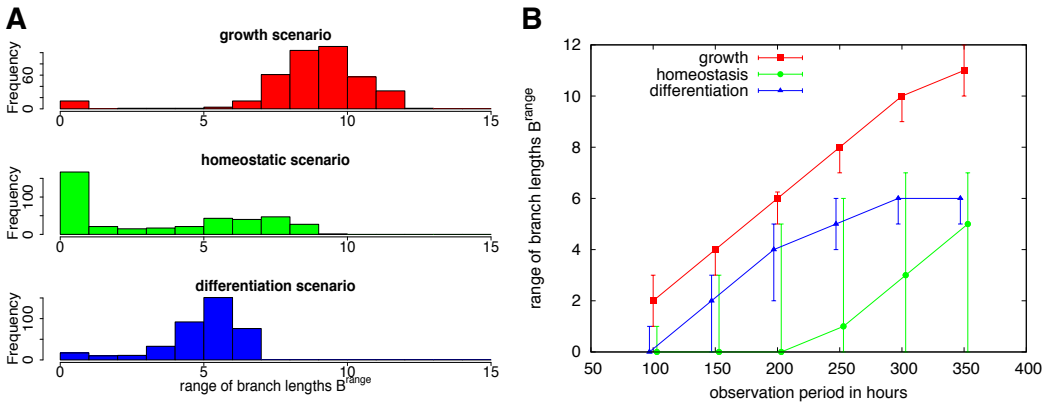

Figure 7.4.: Range of branch lengths B^{range}.
(A) shows histograms of the distributions of the range of branch lengths B^{range} for the three artificial culture scenarios. **(B)** provides the corresponding median values (indicated by the dots and supplemented by the first and third quartiles) for different observation periods (ranging from 100 h to 350 h).

weighted equally, irrespective of whether they occur early or late in development. Therefore, a weighted Colless' index C^w has been introduced in Section 6.3 to account for the exponential expansion within the genealogy putting higher weight on early asymmetries.

As visualized in Figure 7.5(A), there is a tendency for higher values of the weighted Colless' index C^w in the *homeostatic scenario*. It is here that the balanced situation between quiescence and proliferation leads to a number of highly asymmetric genealogies (indicated by high values of C^w). However, at the same time the width of the distribution indicates the appearence of a number of almost symmetric genealogies (i.e. low values of C^w). In these, the branches are committed equally to either continuous proliferation or quiescence. Since cell proliferation is more likely in the *growth* and the *differentiation scenario*, average values of the weighted Colless' index C^w are slightly reduced indicating higher numbers of more symmetric genealogies. Generally, the width of distributions for the weighted Colless' index C^w indicates that the there is a significant population inherent heterogeneity in all simulated scenarios ranging from almost symmetric genealogies to highly asymmetric counterparts.

Figure 7.5(B) illustrates that the weighted Colless' index C^w does not depend on the observation period and, thus, resembles an invariant measure of imbalance in cellular genealogies. In contrast, this scaling invariance does not apply to the classical Colless' index C (data not shown).

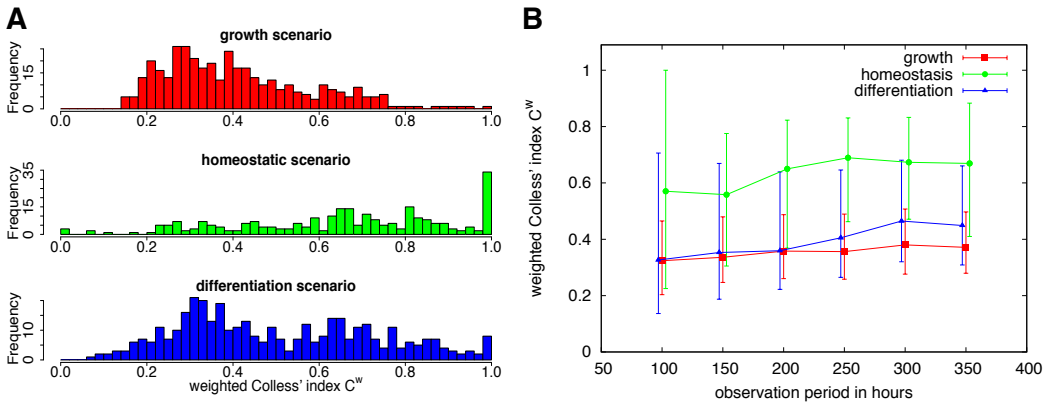

Figure 7.5.: Weighted Colless' index C^w.
(A) shows histograms of the distributions of the weighted Colless' index C^w for the three different scenarios *growth, homeostasis* and *differentiation*. **(B)** provides the corresponding median values (indicated by dots and supplemented by first and third quartiles) for different observation periods (ranging from 100 h to 350 h).

Cell death index A. The cell death index A^g is introduced to estimate the probability for the occurrence of a cell death event conditioned on the particular generation g. This measure can be used to account for changes of the particular probability during the course of differentiation.

Unlike in the experimental situation, in which the role of cell death / apoptosis potentially changes in the course of differentiation, the random occurrence of the background cell death in the simulation model makes a distinction for different generations g obsolete. For simplicity, a generalized cell death index A averages over all generation-depended values A^g for each genealogy (except the root cell generation). Histograms of the corresponding distributions are shown in Figure 7.6(A). Due to the increased proliferation activation and the resulting shortening of G1 phases in the *growth scenario*, the cell death index A is reduced compared to the other scenarios.

Within the simulated scenarios, cell death occurs with intensity $\Phi^B = \phi = 0.02$ at each time step in the simulation model. For a typical G1 phase of 12 hours (corresponding to 12 time steps), the cumulative probability to encounter a cell death event within one cell cycle is calculated as $\Phi_{G1}^{\text{cum}} = (1-0.98^{12}) = 0.215$. This value is well approximated by the generalized cell death index A measured from the simulated genealogies. As shown in Figure 7.6(B), the index values for the *homeostatic* and for the *differentiation scenario* converge towards this analytical estimate for sufficiently long observation periods. Lower index values for the *growth scenario* are plausible, since shortened cell-cycle times (by shortening of G1-phase, see [47]) reduce the probability of induced cell death.

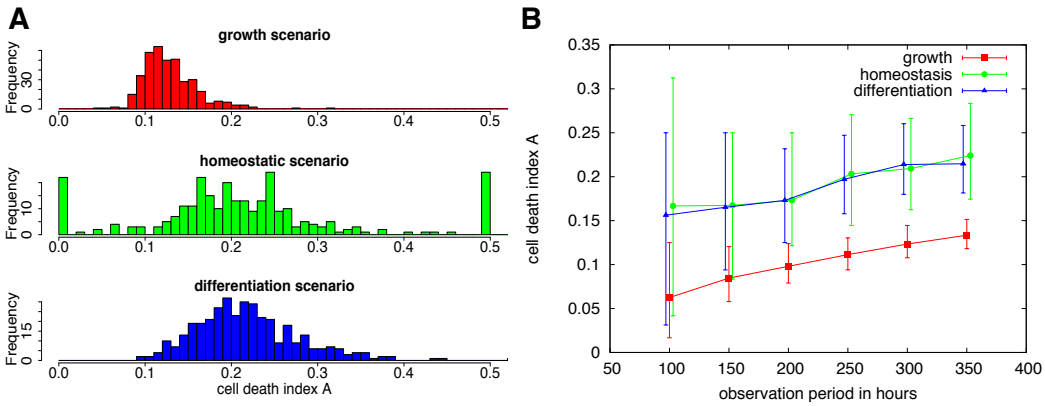

Figure 7.6.: Cell death index A.
In **(A)** three histograms are shown for the distributions of the cell death index A in the *growth, homeostasis* and *differentiation scenario*. **(B)** provides the corresponding median values (indicated by the dots and supplemented by the first and third quartiles) for different observation periods (ranging from 100 h to 350 h).

Minimal distance between characteristic events R. The minimal topological distance between characteristic events R is a measure to quantify whether such events appear correlated between closely related cells or whether they occur isolated within the cellular genealogy.

Histograms in Figure 7.7(A) show the distribution of the average minimal distances R in the three relevant model scenarios. By definition, genealogies with less than two cell death events are excluded from the calculation of the minimal distance measure R. Generally, the minimal distances R between cell death events are rather similar for the three different model scenarios due to the underlying assumption of randomly occurring cell death events acting identically in all three scenarios. However, since cell death events are less likely in the *growth scenario* (compare cell death index A in Figure 7.6(A)), the median of the average minimal distances R is slightly increased compared to the other two scenarios.

For the simulated culture scenarios, in which background cell death occurs randomly with intensity $\Phi^B = 0.02$ during G1 phase, the minimal distance R appears to stabilize around $R = 2$ for sufficiently long observation periods. This behavior is illustrated in Figure 7.7(B).

Briefly summarizing, the application of the proposed measures to three generic model scenarios illustrates their potential for the application to experimental data. In the first place, averages over the single cell measures (e.g. the number of leaves L and the characteristic branch length B^{char}) sufficiently reproduce overall population data (e.g. on cellular expansion). However, beyond these population

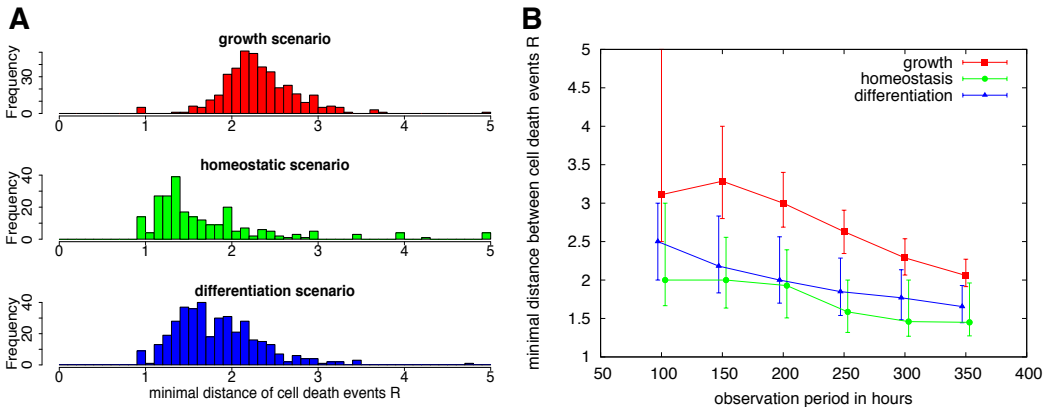

Figure 7.7.: Minimal distance between characteristic events R.
In **(A)** histograms are provided for the distributions of the minimal distance between characteristic events R for the three artificial culture scenarios. **(B)** provides the median values of R (indicated by the dots and supplemented by the first and third quartiles) for different observation periods (ranging from 100 h to 350 h).

averages, the distribution of the measures for a set of genealogies provide an essentially novel access to estimate the population inherent heterogeneity. As a second advantage, the single cell analysis allows to evaluate certain characteristic events within the divisional context (e.g. occurrence of cell death, changes in morphology, lineages decisions). As illustrated above, the generation and the time point of the occurrence of characteristic cell death events as well as the relation to similar events occurring in other cells can only be quantified on the single cell level. Although the proposed list of measures might not be complete and requires adaption for certain experimentally motivated questions it is a basis for the evaluation and statistical comparison of larger numbers of cellular genealogies.

7.1.3. Lineage fate assignment

Two different modes of lineage assignment in cellular genealogies, namely the *prospective* and the *retrospective* view, have been introduced in Section 6.4. Comparing the cellular genealogies it appears that cells at certain positions are already marked as committed in the *retrospective* view, while the *prospective* view indicates that the lineage specification process has not reached a detectable threshold. This notion is supported by a typical cellular genealogy of the *differentiation scenario* which is shown both in *prospective* and the *retrospective* view in Figure 7.8(A) and (B), respectively.

In the context of the simulated culture scenarios it is now possible to quantify this effect. For statistical evaluation, the occurrence of symmetric, asymmetric or

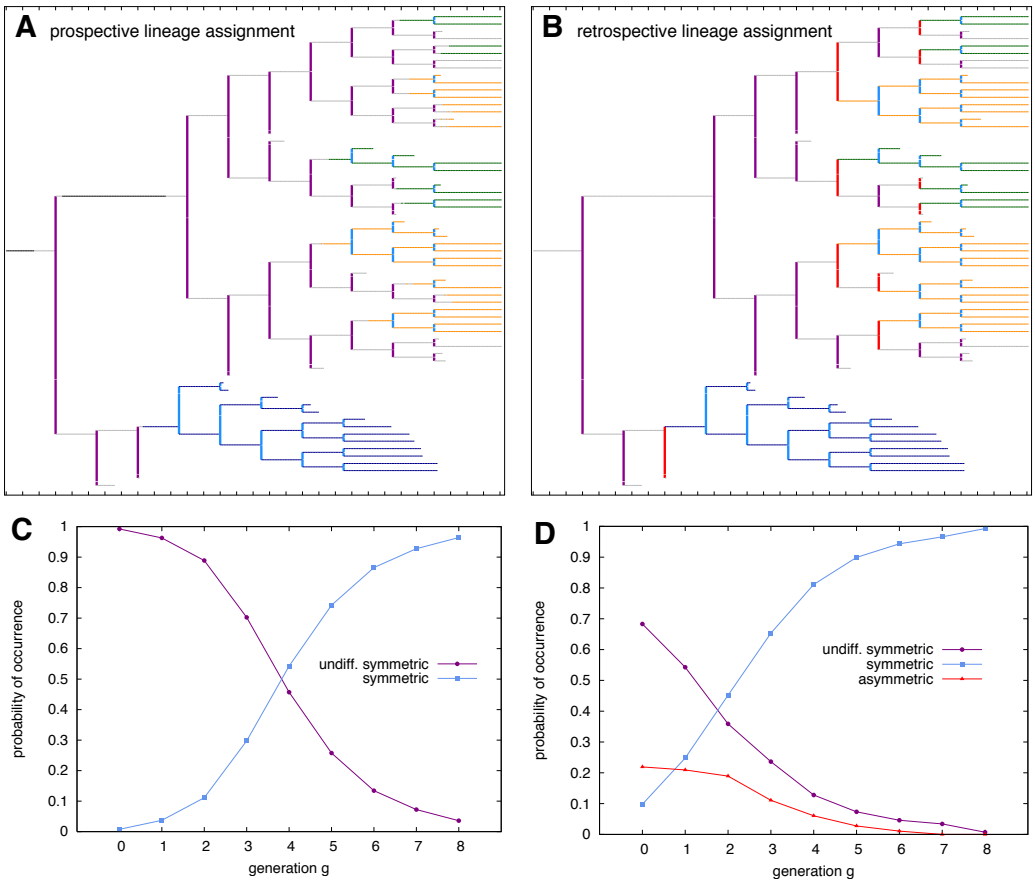

Figure 7.8.: Prospective versus retrospective view for the lineage assignment.
(A). Lineage fate is assigned in situ for a chosen cellular genealogy of the differentiation scenario in the *prospective* view, e.g. the "color coding" of a cell might change during the cells existence if certain critical markers (lineage propensities in the simulation model) exceed a threshold level (i.e. $x_i^* > x_{\mathrm{com}}$). In contrast, in **(B)** lineage fate is assigned recursively based on the lineage fate of the daughter cells for the same genealogy as in (A). Color-coding of the cells is identical to Figure 7.1 (neglecting the early committed stages), colors for the divisions are assigned as follows: undifferentiated symmetric divisions - magenta, symmetric divisions of committed cells - light blue, asymmetric divisions (only in the retrospective view) - red. In **(C)** and **(D)** the probability of occurrence of the particular division types for each generation g is given for the set of genealogies derived under the differentiation scenario. The color-coding is identical to (A) and (B).

undifferentiated symmetric division events, as defined in Section 6.4, is summarized in Figure 7.8(C) and (D). Starting from a population of rather uncommitted cells such histograms are plotted against the generation g in which the division event occurs. Although both fate assignments are based on the same set of underlying genealogies, in the *prospective* view (Figure 7.8(C)) symmetric expansion of undifferentiated cells in early generations (shown in magenta) is more pronounced compared to the *retrospective* view (Figure 7.8(D)). It is the particular construction of the lineage assignment in the *retrospective* view (based on subsequent cellular development and decoupled from the actual intracellular differentiation state) suggesting an earlier onset of lineage commitment compared to the *prospective* view. Although the propensity of a cell for the development in one particular fate might already be skewed at such an early time point, the *prospective* view indicates that fixation is not yet accomplished. This bias is inherently present in any retrospective assignment of cellular characteristics and marks a central disadvantage to the *prospective* view in which critical steps of the lineage specification process are determined in their divisional context.

However, the *retrospective* view is a helpful tool to identify cells that give rise to more than one lineage fate (multipotent cells). Although the multipotency is not based on the transcriptional state of the cell but on its future development, the retrospective lineage assignment is well suited to detect the occurrence and timing of division events that give rise to different (asymmetric) cell fates. In this respect the *retrospective* view illustrates the difference between a functionally asymmetric division, which does by construction not occur in the underlying model system, and an asymmetric cell fate, which is commonly detected in the resulting genealogies.

7.2. Comparing cellular genealogies under different modes of lineage specification

In Section 7.1 the question has been addressed how different scenarios for cellular development influence the shape of the resulting genealogies. Beyond a refined analysis of the population behavior the cellular genealogies can provide additional information about the underlying mechanisms of developmental process leading to the particular topological patterns. The idea is similar to the analysis of phylogenetic trees in which certain apparent features can indicate periods of fierce competition and extinction of species during evolution.

In the following the prevalent debate about the *instructive* versus *selective* modes of lineage specification in hematopoietic stem cells (see also Section 2.3.2) is used as an example to illustrate the potential of cellular genealogies to reveal underlying mechanisms of their generation. Briefly summarizing, the question centers around whether lineage specification is an *instructive* process in which

a combination of cytokines and cell fate specific signals influences the undifferentiated cells such that certain lineages are promoted whereas others are not or whether lineage specification is *selective* in the sense that the intrinsic influence on the cells is negligible, but the regulation occurs on the level of differential survival signals. In the latter setting, cytokines promote the survival of certain lineages whereas cells determined towards other lineages are not supported and consequently undergo cell death [94, 124, 231].

Experimental approaches based on cell population averages are in most cases insufficient to answer the outlined questions for two reasons: first, stem cell populations have a certain, hardly reducible degree of inherent heterogeneity which makes it extremely difficult to initiate cultures of identical and synchronized cells. Second, the population approaches do not capture the temporal evolution and chronology of cellular development as it occurs within a single cell. However, it is precisely the development of each individual cell and its progeny that represents a possible realization of the developmental sequence and retains much of the necessary information: on the correlations between differentiation and cell cycle regulation, on the timing of lineage specification processes and cell death events as well as on the role of asymmetric developments.

The comprehensive mathematical model of HSC development introduced in the previous chapters forms the background to study *instructive* and *selective* lineage specification in an idealized and well defined environment. The understanding of the principle effects of these different modes of lineage specification on the population but also on the single cell level are the prerequisite for the interpretation of data that hopefully becomes available in the next decade.

7.2.1. Generation of cellular genealogies under different modes of lineage specification

For the scope of this analysis a population of bipotent cells is considered, i.e. each cell can differentiate in either of two possible lineages ($i = 1, 2$; shown in blue and green in Figure 7.9). Subsequently, two different modes of lineage specification are applied, both favoring (without loss of generality) the development of the "green" lineage ($i = 2$). Briefly recalling the formal introduction in 4.3.4, in the *selective* mode it is assume that the cell-intrinsic commitment process is balanced and promotes the development of both possible lineages alike ("balanced" rewards $m_1 = m_2$). However, there is a selective cell death process (mediated e.g. by the abstract Pro'green'-cytokine) preferentially affecting cells that initiated development towards the suppressed lineage ($i = 1$) whereas the preferred lineage ($i = 2$) is largely unaffected ($\phi_1 > \phi_2 \approx 0$). In contrast, in the *instructive* mode the cell-intrinsic commitment process is biased towards the preferred lineage ("unbalanced" rewards $m_1 < m_2$), potentially mediated by the abstract Pro'green'-cytokine. In this scenario, cell death occurs randomly in all cells (back-

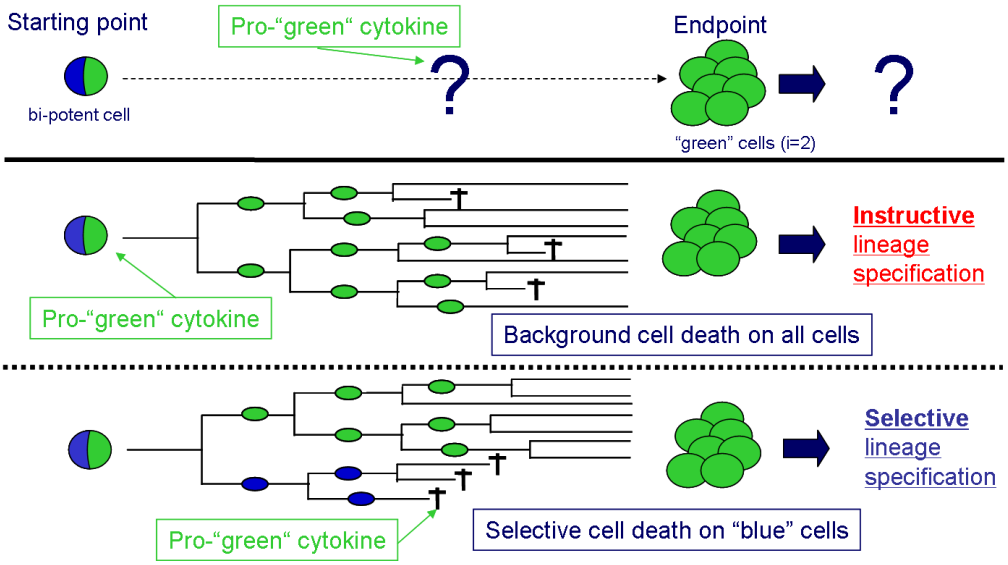

Figure 7.9.: Instructive vs selective lineage specification for bipotent cells. In a general cell culture scenario (shown on top) the potential action of a Pro-'green' cytokine results in the formation of a population of "green" cells. The mechanisms are only accessible on the single cell level: in the case of *instructive* lineage specification the Pro-'green' cytokine intrinsically propagates commitment to only the "green" lineage while background cell death acts randomly on all cells. In contrast, in the *selective* lineage specification both the "green" and "blue" lineages are promoted equally, however selective cell death primarily targets the "blue" cells.

ground cell death Φ^B). The principle idea is outlined in the sketch in Figure 7.9.

For the particular simulations two cell populations of 500 uncommitted, bipotent cells (initial lineage propensities $x_1 = x_2 = 0.5$) with impaired self-renewal ability (regeneration rate $r = 1.0$) are initialized and undergo lineage specification in either the *instructive* or *selective* mode. The tracking process for each of the genealogies extends over 200 hours, thus generating 500 representative cellular genealogies for each scenario. Further parameters are provided in Appendix D.8.

To answer the question whether a statistical analysis of the cellular genealogies reveal any additional information that is not contained in the population kinetics, the parameter configuration is chosen such that the population kinetics (exponential expansion and fraction of lineage committed cells) are indistinguishable for both scenarios. Figure 7.10 illustrates a comparison of the population characteristics.

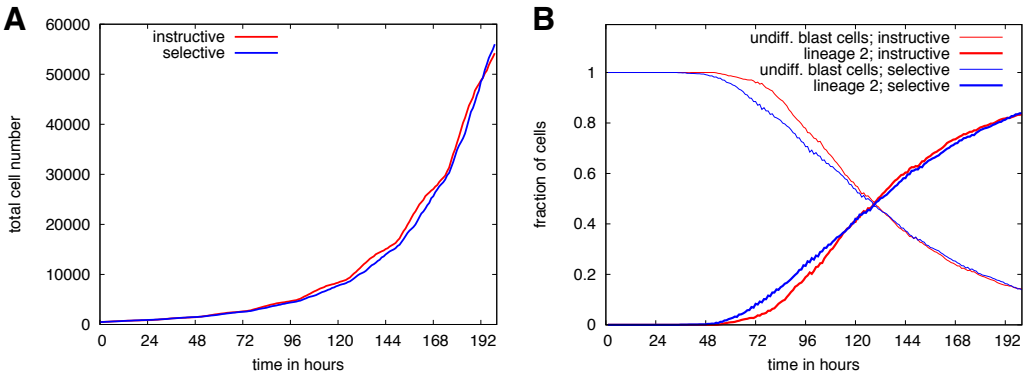

Figure 7.10.: Population development.
The growth kinetics (in terms of the absolute cell numbers, **(A)**) and the temporal development of the lineage specification (in terms of the fraction of uncommitted and committed cells, **(B)**) are shown for both cell populations (500 initial cells in either the *instructive* (red) or *selective* (blue) mode of lineage specification). In (B) the decline of undifferentiated blast cells and the appearance of committed cells of generic type $i = 2$ are provided.

7.2.2. Analysis of structural differences in the topologies

The measures introduced in Section 6.3 are applied to address the structural differences in the shapes of the genealogies generated by use of either the *instructive* or the *selective* mode of lineage specification.

Both, the total number of leaves L and the characteristic path lengths B^{char} are constructed to quantify clonal expansion. As one example the histogram of the distribution of the number of leaves L is provided for both modes of lineage specification in Figure 7.11(A). Since the population kinetics in Figure 7.10 have been fitted to resemble almost identical overall growth behavior of the cell cultures, these findings are reflected on the single cell basis, too. Neither the mean value nor the shape of the distribution show pronounced differences that could possibly be assessed by a limited number of realizations.

As a further example the histogram for the weighted Colless' index C^w is shown in 7.11(B). One could speculate that the action of selective cell death as compared to random (background) cell death introduces asymmetric genealogies in which one branch expands more than the other. Colless' index C^w is designed to detect such asymmetries in tree-like structures, however there is no considerable difference to be detected between the *instructive* and the *selective* mode of lineage specification. Although the shapes of the distributions between both scenarios are somewhat different, a quantification of this effect is unlikely to be measurable.

In Section 6.3 a number of measures has been introduced addressing the mutual relation between characteristic events, namely the minimal distance between

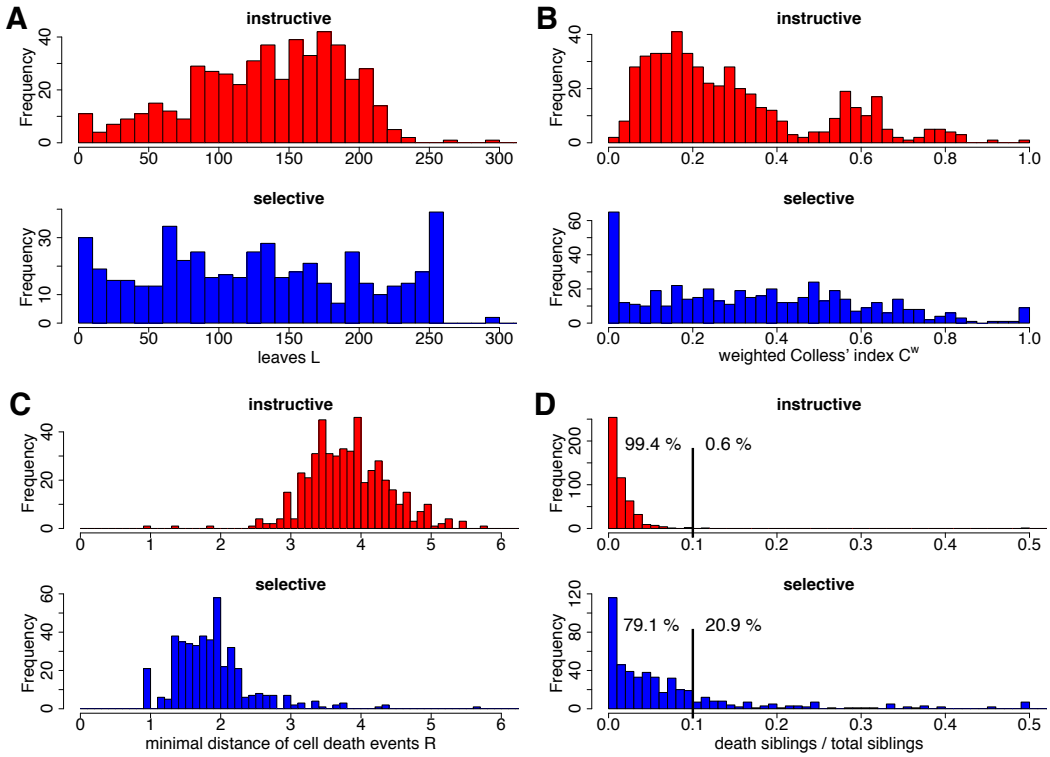

Figure 7.11.: Selected measures for cellular genealogies.
The histograms compare the cellular genealogies derived from 500 initial (root) cells in either the *instructive* (red) or *selective* (blue) mode of lineage specification for the following selected measures: **(A)** total number of leaves L (shown on a logarithmic scale), **(B)** weighted Colless' index C^w, **(C)** minimal distance between cell death events R, **(D)** fraction of death siblings.

characteristic events R and the mutual information measure MI. In the *selective* scenario cells undergo cell death in case they are developing in the unfavorable lineage. As similar developments are more likely in closer related cells (by sharing a common ancestry) selective cell death is more likely to appear in close relatives. The histogram in Figure 7.11(C) for the minimal distance between cell death events R supports this notion. Although the shape of the distributions is similar, there is a pronounced shift in the mean values of R.

A similar observation can be made using a simplified version of the mutual information measure MI. Figure 7.11(D) shows a histogram for the normalized fraction of sibling pairs (two cells directly derived from on common parental cells) in which both cells undergo cell death before they can initiate a further cell division ($c_p, c_q \in \mathcal{C}^{\text{death}}$; c_p, c_q are siblings). In both scenarios, there is a rather large fraction of genealogies in which cell death of both siblings does not occur at

all. However, in the *selective* mode of lineage specification there is a considerable fraction of genealogies in which this "harmonized cell death" occurs far more frequently as compared to the *instructive* case. The numbers within the histograms indicate the percentage of cellular genealogies in which the death of both sibling cells occurs in less/more than 10 % of the observed sibling pairs: whereas this is an extremely rare event under the assumption of an *instructive* lineage specification (less than 1 % of the cellular genealogies) it is far more likely under the *selective* mode of lineage specification ($\approx 21\%$ of the observed cellular genealogies). The mutual information MI is less suited to detect this "heavy tail" as it is more focused on the average values. Therefore, it is not shown for the particular case.

Briefly summarizing, it appears that in a situation in which two expanding cell populations are indistinguishable on the population level, appropriate measures on the single cell level can still detect significant differences. In particular the measures focusing on the occurrence and correlation of characteristic events proved especially useful.

7.3. Summary and additional remarks

Using the presented simulation model of HSC organization it could be shown that cellular genealogies contain a number of additional information which is not accessible on the population level. As for the example on *instructive* versus *selective* lineage specification it has been demonstrated that cellular genealogies reveal "footprints" of the underlying decision process. An additional and important aspect making the analysis of cellular genealogies a highly rewarding technique is the direct access to the population inherent heterogeneity. As indicated by the model studies in Section 7.1 a high level of heterogeneity (e.g. with respect to cellular expansion) is expected even among a population of closely similar cells. It will be a central task of future research in cell biology to identify the mechanisms that establish and maintain these population inherent heterogeneity and to explain their functional relevance. The recording and analysis of cellular genealogies is one promising technology in this direction.

However, there are currently a number of technical limitations to fully exploit the information contained in extended cellular genealogies. Cutting a long story short, it is still a major experimental and bioinformatical challenge to generate sufficiently long time lapse videos at reasonable cell densities and to extract enough reliable cellular genealogies from them. Although one genealogy might be enough to document the existence of a particular common progenitor (as shown in [225]) the statistical comparison between different sets of genealogies requires significantly higher trial numbers. Apart from these technical limitations there is

a further structural difficulty. Studying a particular experimentally derived set of cellular genealogies, the application of the outlined measures does not ultimately allow the identification of the particular mechanism (e.g. with respect to the lineage specification to be either *instructive* or *selective*) since the necessary reference scenario is generally missing. In turn, this is a strong argument in favor of the presented theoretical approach. It has been shown that the mathematical model can be adapted to the population kinetics for the cell culture in question using either the *instructive* or the *selective* mode of lineage specification. The resulting genealogies can then act as the reference scenarios to which the experimental data is finally compared.

One could also argue that the estimation of an average cellular genealogy combining the characteristic features of a large set of cellular genealogies might be a good representation of the underlying development. However, applying such a method all the information about population inherent heterogeneity is abandoned. Furthermore it has been shown for the distinction between *instructive* and *selective* lineage specification that not only the mean value but especially the outliers and the frequency of their occurrence are better predictors. This is especially visible for the case of the fraction of death siblings (Figure 7.11(D)) in which the frequency of genealogies with high values is distinctly increased for the *selective* lineage specification.

The results presented in Sections 7.1 and 7.2 are published in [125, 232].

8. Discussion and conclusion

8.1. Summary of results

The mathematical model presented within this thesis is the first to describe the individual commitment process of single cells within a whole population of hematopoietic stem cells and their progeny. It represents an **essentially novel, systems biological approach to the quantitative elucidation of lineage specification as a temporally extended, self-organizing process**. The model has been successfully applied to reflect different experimental situations. Furthermore it could be demonstrated using this model that the tracing of individual cells and their progeny contains additional information which is not available based on population studies alone.

The discussion of the results within this chapter follows the outline of the thesis. Section 8.1.1 provides a critical assessment of the lineage specification model and its implementation within the pre-existing model of HSC self-renewal. The application of the model for the analysis of single cells and their clonal offspring is discussed in Section 8.1.2. A comprehensive conclusion and outlook are provided in Sections 8.2 and 8.3.

8.1.1. Amended model of HSC self-renewal and lineage specification

The description of lineage specification in terms of intracellular propensities represents an abstracted, *phenomenological perspective* on the underlying molecular dynamics within one individual cell. In order to understand the consequences of this intracellular regulation on the tissue level, the model of lineage specification is consequentially integrated into a more general, single-cell based model to describe HSC organization originally proposed by Roeder and Loeffler (see Section 4.4). This comprehensive, amended model of HSC self-renewal and lineage specification is the basis upon which different aspects of HSC organization are reflected.

The presented model concept is discussed with respect to the criteria stated in Section 4.1:

- **Lineage commitment (C1) and generation of diversity (C2).** The model supports the idea of a progressive restriction in lineage potential in the course of differentiation, i.e. starting from a population of multipotent cells the potential to contribute to all possible lineages declines in the course of differentiation. To make this idea more visible the lineage contribution of

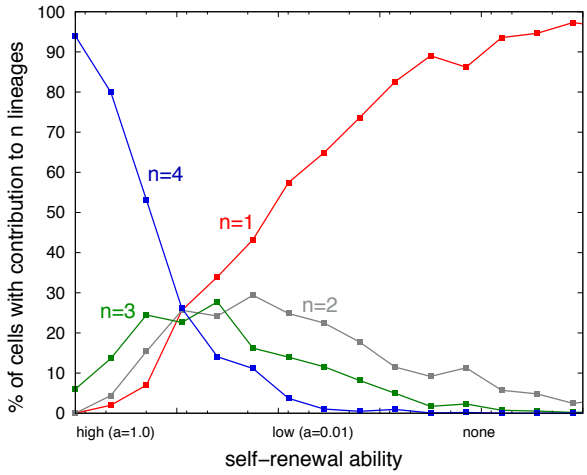

Figure 8.1.: Lineage contribution versus self-renewal ability.
The curves correspond to the fraction of cells that contribute to n out of $N = 4$ possible different lineages within their differentiating progeny. The curves are shown as functions of the initial self-renewal ability of the parental cell measured by the affinity a in the model simulations. 50,000 simulated, individual cells are randomly taken from a homeostatic system (with their particular affinity a) and put into an empty model system in which they undergo lineage specification. The number of observed lineages n is counted for the progeny of each individual cell.

single differentiating cells is illustrated depending on their initial repopulation ability. In terms of the model the repopulation ability is characterized by the affinity parameter a. For the particular example of a system with four possible lineages the actual contribution of a cell's progeny to n out of the $N = 4$ available lineages is shown as a function of the affinity parameter a (Figure 8.1). As expected, nearly all cells with high repopulation ability ($a \approx 1$) contribute to all four lineages, while tri- and bipotent cells are found mostly in the population with lower self-renewal ability (i.e. $a \approx a_{\min}$). The ultimate loss of repopulation potential is associated with commitment to a single lineage.

It is essential to note that the hierarchic decrease of lineage potential is not due to a cell-intrinsic, pre-defined developmental program, but emerges as a system-inherent property of the proposed model. Therefore it follows that the hierarchical appearance of the lineage specification process can be consistently embedded in the context of fluctuating, self-organizing and flexible stem cell populations.

Another important aspect is the distinction between *lineage contribution* (being the lineages actually generated by the progeny of a particular cell)

and *lineage potential* (being the lineages to which the same progeny could have contributed). Since a single cell can only differentiate once, the lineage potential can not be determined experimentally. However, despite this inherent uncertainty, the notion of lineage potential is important to understand the organizational principles of cell populations and tissues as well as for the characterization of the stem cell properties. As it has been shown exemplarily for the lineage specification kinetics of the FDCP-mix cell population, the fluctuations in lineage potential occurring on the single cell level average out on the population level. This means that although the outcome on the population level is robust the particular fate of a single cell can only be predicted in a probabilistic sense. Based on this understanding, the proposed model predicts that heterogeneity of a progenitor population is inherently generated as a consequence of the autonomous development of individual cells.

- **Temporal extension and reversibility (C3).** The identification of a broad range of hematopoietic cell types (through which a differentiating HSC passes until its final maturation) suggests that hematopoietic lineage specification is not an immediate decision but at least a stepwise process with a certain temporal extension. The proposed model approach goes one step further by assuming that lineage specification is an intrinsically continuous process. Describing the state of a cell in terms of its intracellular concentrations of molecular components (the concept of "state space", compare Section 3.3.2 in which the components represent e.g. mRNAs, TF, proteins, receptors etc.) it is rather unlikely that the concentration of any of these components changes in a discontinuous manner. Under the necessary assumption that the phenotype of a cell corresponds to its molecular circuitry, this perspective ultimately implies a continuous approach to any type of cellular development.

  Technically, lineage specification is described by lineage propensities x_i taking any value between 0 and 1. In contrast, the update procedure operates on discrete time steps such that the underlying continuous process can only be approximated by the numerical model. However, as long as the time steps are sufficiently short this approximation does not alter the system dynamics. This criterion is generally fulfilled for the presented model simulations. More critically, the phenotypic mapping introduced in Section 4.3.3 imposes a rather arbitrary classification between "uncommitted" and "committed" cells. However, the threshold values x_{com} do not influence the underlying competition process and are solely introduced for the purpose of comparison with experimental results. In this respect the choice of x_{com} reflects an additional degree of freedom for the fitting procedure.

The perception of lineage specification as a continuous, temporally extended and fluctuating process integrates into the functional definition of stem cells. The common understanding of HSC organization as a well-structured, discrete hierarchy (outlined in Figure 2.2) only appears as a snapshot of an underlying, continuous transition of cells from the multipotent stem cell state towards the more differentiated cell states. In this respect, the transition from a multipotent stem cell over a lineage restricted progenitor towards a mature, functional cell is considered to be gradual instead of discrete.

The assumed temporal extension of losing multi-lineage potential is closely associated with reversibility of the differentiation process. Lineage specification is modeled as a process favoring certain lineage developments by progressively decreasing the probabilities for the competing options. Therefore, reversibility depends strongly on both, the actual state of differentiation and on the influence of the micro-environment. The model predicts that reversibility of lineage specification is a rare event in a homeostatic system but occurs more frequently in the disturbed situation with the need for system repopulation. This finding is in good qualitative agreement with experimental and clinical observations suggesting an increased flexibility of stem cells in the situation of tissue repair. This effect, commonly referred to as "plasticity" is mostly used if stem cells of a particular tissue contribute to another, rather unrelated tissue in the case of high demand (e.g. bone marrow cells contributing to regeneration of muscle [20, 21] or infarcted myocardium [15]). However, plasticity also refers to an increased flexibility of tissue stem cells to reversibly change between different developmental options within their particular tissue [33, 96].

In the experimental situation similar effects should be observable when cells that are primarily cultured in a particular differentiation promoting medium are subsequently transferred into a condition with different properties (e.g. promoting self-renewal or another differentiation program). The model predicts that the fraction of cells with "reverted" development is not an all-or-nothing decision, but depends in the first place on the exposure time in the particular medium. A rigorous experimental test of this prediction would have to use molecular markers that are irreversibly switched on if a certain characteristic gene expression identifies a particular lineage commitment. The detection of such markers can elucidate to what extend early committed cells actually "reverse" their previous development under changing environmental conditions. The model predicts that the fraction of cells with reversible developments gradually decreases as the process of lineage specification continues.

- **Regulation of lineage specification (C4).** The question whether particular lineage decisions are governed by instructive or selective mechanisms is long standing (cf. Section 2.3.2). As shown in Section 4.3.4 the model is able to reflect both these aspects. Whereas for the *instructive* lineage specification the lineage specific rewards m_i are adjusted, a lineage specific cell death Φ^S is assumed for the *selective* scenario.

  It is demonstrated in Section 5.4.3 that for a particular experimental situation both scenarios of lineage specification are equally suited to describe the observed results. Based on the analysis of the extended model system in Section 5.2.3 it is most likely that this equivalence applies to many experimental datasets in which lineage specification is described on the population rather than on the single cell level. However, it has been demonstrated within this thesis that a distinction between *instructive* and *selective* lineage specification can possibly be made based on the analysis of single-cell developments (also referred to as *cellular genealogies*). It is a remarkable feature that in case of *selective* lineage specification cell death events are more likely among closely related cells (such as siblings) as compared to the *instructive* scenario. Such effects can be detected using appropriate methods for the analysis of cellular genealogies as they are suggested in Section 7.2.

- **Priming as a molecular concept of multipotency (C5).** The model concept supports the hypothesis that the experimentally observed *priming* behavior (i.e. the low level coexpression of lineage specific and potentially antagonistic transcription factors) is a common molecular representation of the stem cell state (see also Section 2.3.2 and references [59, 43, 233]) which is maintained under specific conditions (e.g. due to niche signals). Maintenance of the priming state could feasibly be achieved by active epigenetic stabilization of chromatin structures retaining parallel developmental options. Changing micro-environmental signals might be one possible mechanism to destabilize the priming state. Under these modified conditions, chromatin changes at key loci may result in a sequential shift of the expression state towards one or the other lineage specific expression pattern. At an experimentally accessible level, the model predicts that targeted up- or down-regulation of certain lineage specific genes upsets the balance at the priming level and consequentially supports or discriminates certain options in the subsequent differentiation process. The model furthermore suggests that such processes are best studied on the level of single cells. Alongside with the experimental tracing of individual cells in culture it is shown in Chapters 6 and 7 that the model is able to mimic critical phenomena of the differentiation sequence on the single cell level and to couple them to the population level.

Placement of the model. The presented model is in line with similar concepts perceiving lineage specification as a progressive restriction in lineage potential [26, 218]. Although the decision process is intrinsically random it is tunable by amending the lineage specific rewards m_i. The model was intentionally constructed from a *phenomenological perspective* with a focus on the maintenance and loss of multipotency and the temporal extension of the commitment process. This approach is structurally different from the *functional perspective* (cf. Section 3.3.4) describing the interaction and sequential down-regulation of certain lineage specific transcription factors during differentiation as proposed for different small-scale networks [92, 108, 234].

To my knowledge the presented model is the first approach to integrate a description of intracellular lineage specification dynamics within a functional model of HSC organization. In this respect the model couples intracellular decision processes to the phenotypic appearance of the overall tissue system. It is this multiscale description that allows to study different mechanisms of lineage specification (such as *instructive* versus *selective*) on the intracellular level and interrogates the consequences on the tissue level. Based on this philosophy of self-organizing, heterogeneous and mutually interacting cells, similar integrative models have been recently developed for mesenchymal stem cells [235] and for stem cells of the intestinal crypt (Galle, Loeffler et al., in preparation).

The concept of "asymmetric cell division" has been used regularly to describe the balance between self-maintenance of a stem cell population and the differentiation into tissue cells (compare Section 3.2.1). Although asymmetric cell division have been reported for a number of stem cells systems [185, 186] it is not yet clear whether such events do also occur in the hematopoietic system and whether they are functional with respect to the asymmetry occurring in the development of the daughter cells and their progeny. Therefore, the concept of functionally asymmetric divisions in hematopoietic stem and progenitor cells is still an unproven hypothesis. These doubts are strengthened by the presented model demonstrating that a consistent explanation of the heterogeneity among differentiated cells on the tissue level is possible without assuming an asymmetric division process on the intracellular level. Within the proposed model any simulated cell division is symmetric by definition. Differences in the individual development of the daughter cells occur only due to their independent differentiation sequences after mitosis. Asymmetric development is thus interpreted as the asymmetry of cellular fates, not of the division process itself [183]. More technically speaking, asymmetric fate is solely the result of the independent development of the two daughter cells after a functional symmetric division event.

Limitations of the modeling approach. The particular underlying mathematical process of the lineage specification dynamic was chosen because it re-

sembles the criteria outlined in Section 4.1. However, the process is based on a number of simplifying assumptions limiting the transferability to a directly measurable molecular process. For instance, lineage specification dynamics presumably require a set of many coregulated factors which have been summarized into one generic lineage propensity. This simplification neglects subsequent activation steps, mutual interactions between the members of each of the sets of coregulated factors, and the role of late signaling events. Similarly, the role of extrinsic signaling by cell-cell and cell-environment interactions is reduced to the influence of two antagonistic control regimes governing the lineage specification process. Furthermore, the phenotypic mapping to classify cells as either uncommitted or committed (introduced in Section 4.3.3) is only a rough approximation of the highly complex maturation process. Despite, or perhaps because of this simplicity, the model proves sufficient to account for a considerable number of phenomena on the lineage specification of hematopoietic stem cells. Most notably, all these results are consistent with previous findings on self-renewal and clonal competition (see also the summary in Table 8.1).

This is not to say that the provided explanation is either unique or complete. Indeed, a detailed quantitative understanding of lineage specification must eventually take account of the characteristics and interactions of a plethora of regulatory molecules to cover the full complexity of intracellular regulations. However, as outlined in Section 3.3.4 such an approach requires a better and more detailed understanding of the involved molecular components and their mutual interactions.

The interpretation of the data analysis presented in Section 5.4 requires a critical acclaim. For the simulation of the experiments by Suda et al. [26, 27] and Takano et al. [222] the particular cell sources are defined as a certain subpopulations of the pool of simulated cells, referred to as “root pools”. These are adjusted to represent the balance between multipotent cells and committed cells (compare also Figure 8.1 in which the decline of multipotency as a function of the attachment affinity a is illustrated). However, any developmental preferences for certain lineages or correlations between them are neglected and there is no adaptation of turnover rates which might actually depend on the lineage commitment.

The choice of parameters for the simulations of the lineage specification in differentiating FDCP-mix cell cultures in Section 5.4.3 has been adapted to meet the time scales of lineage commitment as well as the fraction of cells within the individual lineages. The good correspondence with the experimental results is taken to support the idea that lineage specification is tunable on a probabilistic level in which the fate of a single cell can be influenced although a prediction about the future development cannot be made with certainty. However, this uncertainty disappears on the population level as the individual realizations contribute to a dynamically stable “average tissue”.

8.1.2. Tracing the ancestry of single cells

The application of the extended stem cell model to different experimental settings in Chapter 5 suggests that the analysis of single cell fates is more instructive as population averages alone. This thesis provides an essentially novel and innovative approach to analyze and compare such pedigrees of cellular development originating from one particular cell. Based on the tracing of cellular ancestry using time-lapse video monitoring and appropriate tracking software the classical population based approach is abandoned and it becomes possible to perceive the development of each entity separately and to place it within the population context. The resulting cellular genealogies are a rich source of information: on ancestry, division times, morphological changes, motility, cell-cell interactions, and correlations among related cells, just to name a few. However, the extraction of these quantities and their analysis for comparability requires a theoretical foundation.

In this thesis a characterization of the topological features of cellular genealogies is provided alongside with a number of measures particularly addressing the quantitative analysis of individual cell fates including the balance between stem cell proliferation, quiescence, and cell death. Due to limitations in the availability of experimental data, these measures are discussed in the context of the amended model of HSC organization although they are applicable to cellular genealogies derived from different sources.

The introduction and analysis of cellular genealogies in Chapters 6 and 7 revealed a number of prominent aspects:

- **Heterogeneity and the detection of rare events.** The measures for the quantification of cellular genealogies that are proposed in Section 6.3 are primarily suited to distinguish between cellular genealogies derived under different growth scenarios. Additionally, these measures can be applied for the estimation of the inherent variation within a set of genealogies derived under identical conditions. The width of the distributions for the different measures shown in Figures 7.2 to 7.7 illustrates that even the genealogies which are derived under identical conditions show a large variability. In this respect, cellular genealogies and their topological characterizations are powerful tools to quantify clonal heterogeneity, and to distinguish whether stem cell populations are inherently heterogeneous or if they are composed of predefined homogeneous subsets.

  The analysis in Section 7.2 on how *instructive* and *selective* lineage specification influence the topology of the resulting cellular genealogy hints towards another interesting aspect of the single cell analysis. As indicated in Figure 7.11 the occurrence of pairs of death cells among siblings shows a different type of distribution in case of the *selective* lineage specification as compared

to the *instructive* lineage specification. In particular, there is an increased frequency of the cellular genealogies with higher rates of death siblings for the *selective* lineage specification. Although such events are rare they lead to the establishment of a so called "heavy tail distribution" which is not observable for the *instructive* lineage specification. In other words, the distinction between the different modes of lineages specification does not rely on comparison of mean values but is based on the frequency of rare events. This behavior, which is illustrated for one particular example, is one of the reasons why the author avoids the notion of an "average genealogy". It is most likely that the averaging process over many heterogenous genealogies takes away a lot of the relevant information obtained on the single cell level. In this respect, the level of population inherent heterogeneity appears as an equally important classifier of cellular development as compared to the commonly used average values.

- **Occurrence of characteristic events.** The structure of cellular genealogies allows the precise location of different types of characteristic events (e.g. morphological changes, cell death, expression of fluorescence marked reporter genes) within the divisional history of the cell and its "clonal relatives". The hypothesis of whether such characteristic events are related to a cell intrinsic program which advances similarly in closely related cells or whether these events result from external queues can be tested by investigating the mutual relation between the characteristic events. The analysis of Chapter 7 indicates that different measures addressing this relatedness of characteristic events (i.e. whether they occur in closely related cells or only in further distant relatives) are suitable and specific for the quantitative analysis and comparison of cellular genealogies.

  For the illustrated example of cell death occurrence, the functional role of this process for the regulation of lineage composition both *in vivo* and *in vitro* is still controversial [92, 94, 124]). However, if it becomes possible to clearly identify cell death events e.g. by monitoring the activity of certain relevant genes in the apoptosis pathway using fluorescence labeling methods, cellular genealogies are a unique tool to investigate this action in the divisional and in the population context. The proposed measures (i.e. the mutual information MI and the average minimal topological distance between such events R) are possible tools to quantify these aspects.

- **Analysis of cell cycle times.** Apart from the elaborated analysis introduced in Chapters 6 and 7, the availability of cellular genealogies will also allow for an exact characterization of individual cell cycle times τ_c. Whereas classical estimates of cell cycle times are based on measurements of the fold increase in a population of differentiating cells, which neither account for

the heterogeneity of individual cells nor for the occurrence of cell death, the distribution of cell cycle times can be reliably estimated from a sufficiently large set of cellular genealogies. Starting from a paternal division d_i the time interval to the next division d_j is an exact measure of the cell cycle time τ_c. Besides the global distribution of cell cycle times, the representation of clonal development in a cellular genealogy allows the evaluation of cell cycle times with respect to secondary parameters e.g. according to the particular cell generation g or to cell fate specific information that accompany a particular genealogy. In the latter case, correlations between the lineage fate and the change in cell turnover can be quantified circumventing the obstacles of a population average potentially containing different cell types.

Placement of the model approach. The topological characterization and the proposed measures for the analysis of cellular genealogies represent a novel systems-biological methodology for the analysis of single cell developments. Although pedigrees of cellular development have been derived in different experimental situations [124, 126, 130, 225] a statistical analysis of the resulting cellular genealogies, as the one presented within this thesis, is still missing. However, it needs to be remarked that the available data sets are currently not sufficient to apply this methodology with scientific benefit.

The derivation of the cellular genealogies in Chapters 6 and 7 builds upon the comprehensive model of HSC organization introduced in the preceding chapters. It is one central advantage of a mathematical model to allow testing a large variety of possible measures on whether they are suited to identify differences in the generation scenarios. Based on such a strategy a number of such measures could be disqualified which performed poorly in the comparison and characterization of cellular genealogies. Moreover, an *in silico* model can be tuned such to pronounce certain developmental aspects like differentiation at the cost of self-renewal or a bias towards particular lineage fates. Using the predicted model genealogies as a reference system and comparing them with their "real" counterparts (as soon as they become available) is a powerful systems-biological tool to uncover the imprints of different developmental and/or regulatory processes that are hidden in the complex topological structure of this particular type of data.

Limitations of the particular simulation model. The application of a mathematical model represents a unique tool to study the explanatory and the statistical power but also the limitations of certain analysis methods prior to the generation of large amounts of data. However, using a mathematical model instead of biological data bears a number of risks and uncertainties especially for the generalization of the results. Some aspects, that are inherently present in experimentally derived data, can not be studied on the basis of the particular simulation model.

For example, variations of the cell cycle times τ_c are not adequately represented by the model. However, the measures proposed in the Chapter 6 are only based on the topological structure (i.e. the parent-daughter relation) and do not require the estimates of the cell cycle times. Therefore, the drawn conclusions also apply to the more general and experimental relevant situation in which cell divisions are not synchronized.

Furthermore, the simulated genealogies do not account for the migration of cells since the employed stem cell model is not based on an underlying spatial structure. Therefore neither spatial correlations between the existing cells nor their velocities are accessible, and the analysis of their influence on cell fate decisions can not be studied using the current model implementation. The structural characterization of cellular genealogies, as it is presented above, can be easily extended to incorporate the spatial component. A corresponding project is currently under way. Finally, the list of proposed measures is neither complete nor exclusive. Different biological questions might result in the development of novel measures that are particularly designed to reveal certain structures within the genealogies.

8.2. Conclusions

Summarizing the above discussion, there are a number of major conclusions from the presented study. First, is has been shown that the process of lineage specification can be integrated into the previously proposed hematopoietic stem cell model [33, 46] in a conceptually consistent way. The essential idea is a fluctuating, albeit balanced ground state of competing lineage propensities which might be retained or suspended depending on the exposition to appropriate environmental conditions. Furthermore, it has been demonstrated that this combined model of stem cell organization and lineage specification accounts for the phenotypic heterogeneity that is experimentally observed in populations of differentiating stem and progenitor cells and is consistent with the assumption of a progressive restriction in lineage potential. As outlined above, stem cell development and lineage specification are considered as temporally extended and fluctuating processes of continuously changing cellular characteristics. This concept does not exclude certain preferred trends in the differentiation sequence, but it comprises the possibility of reversible developments for individual cells and, thus, allows the system to flexibly react to changing demands. In this sense "stemness" is no longer understood as a cellular feature, but as a system property. This perspective has first been proposed by our group [1, 33] and independently by others [34, 60, 175, 174, 113, 176]. The concept is fundamentally different from approaches describing stem cell organization as the consequence of a predefined, cell-intrinsic differentiation program. Such approaches assume discontinuous transitions from one confined stem cell or progenitor subpopulation to another in a predefined, strictly unidirectional differ-

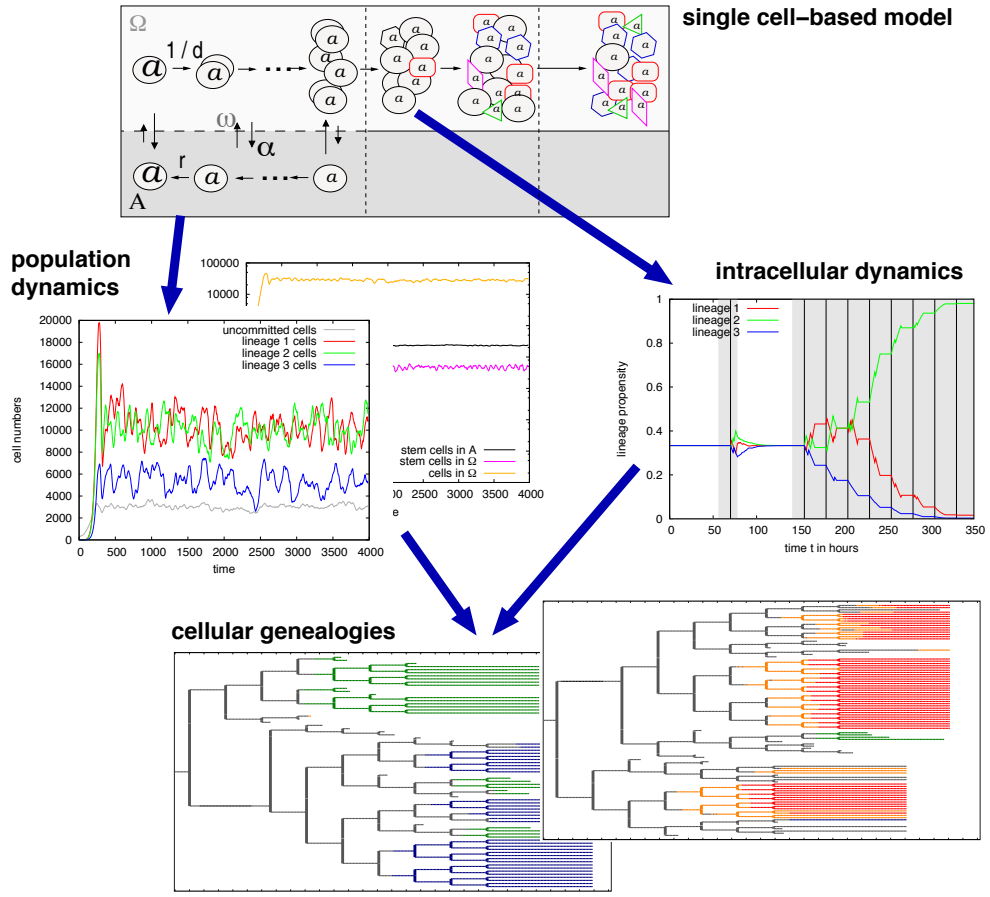

Figure 8.2.: Overview of the extended stem cell model.
The single cell-based model (top) allows to study hematopoietic stem cell development both on the population level (left) and on the intracellular level (right). Aspects of both these levels can be projected conveniently into representations of cellular genealogies (bottom).

entiation sequence [86, 170, 236, 237]. Clearly, the grouping of stem and progenitor cells according to features such as cell surface marker expression and functional characteristics remains useful for classification, selection and enrichment, since it accurately reflects the behavior of a population under a certain set of conditions. Ultimately, however, our increasing awareness of heterogeneity, flexibility, and plasticity within stem and progenitor cell populations questions the validity of these strictly unidirectional concepts at the mechanistic level in single cells.

Second, it has been illustrated that cellular genealogies, derived from the tracking of individual cells over time, are a rich source of information on cellular development that goes far beyond classical population based approaches. Most notably,

it has been demonstrated that, by the introduction of a mathematical characterization, this very complex data type is accessible to statistical evaluation and comparison. It can be expected that the availability of time lapse video microscopy and the establishment of efficient image-processing methods will soon allow the "high throughput" tracing of single cells within cell cultures. The interpretation and management of the resulting cellular genealogies is a challenge to experimental and theoretical biologists alike. Therefore, it can be argued that the development of efficient automized tracking routines on one side but also the establishment of a powerful analysis pipeline on the other side are both integral parts of a joint venture that need to be pursued in parallel.

Taken together, the integration of the intracellular model of lineage specification into the previously described model of self-organizing HSC populations [46, 33] establishes a powerful conceptual *and* numerical tool to address a broad range of experimentally observed phenomena in hematopoiesis. The flow diagram in Figure 8.2 summarizes the different descriptional levels of the model. Starting from the extended, single cell-based model the dynamical behavior is accessible on the population level and on the intracellular level. Cellular genealogies appear as a unified picture accounting for aspects of both these levels.

On the descriptional level the model is consistent with previous findings on stem cell self-renewal and clonal competition. However, the extended approach additionally accounts for the experimentally observed phenotypic heterogeneity in populations of differentiating stem and progenitor cells assayed under various conditions. The overview in Table 8.1 summarizes the coverage of the amended HSC model.

	HSC model by Roeder and Loeffler	amended HSC model including lineage specification
stem cell self-renewal and differentiation	+, covered in [33, 183]	+
clonal competition and leukemia	+, covered in [221, 220]	+
clonal heterogeneity	+, covered in [122]	+
lineage specification		+, covered in [114]
single cell tracking		+, covered in [125, 232]

Table 8.1.: Coverage of the extended HSC model
The table provides an overview of different phenomena in hematopoiesis and their coverage in the original HSC model by Roeder and Loeffler compared to the amended HSC model including lineage specification. Corresponding publications are cited.

8.3. Outlook

The presented thesis is a good example to illustrate two important aspects of theoretical stem cell biology: Whereas the first part concentrates on a mathematical framework to describe the phenomenology of lineage specification in hematopoietic stem cells (*modeling)*, the second part develops a methodology to analyze a particular data type, termed *cellular genealogy*, describing the developmental history of individual cells (*data analysis)*. Beyond the scientific results this work illustrates how experimental results and techniques have a continuation in theoretical biology and how this data analysis and conceptualization feeds back into experimental approaches. It is this combination of mutual stimulus that will make the idea of a Systems Biology of Stem Cells a vivid and productive field of research.

On the practical side, the model approaches of this thesis have different implications for experimental studies:

- The perception of lineage specification as a temporally extended process strongly suggests that cellular development needs to be studied in a time dependent manner. This implies repeated functional, phenotypical and molecular measurements at different time points of a developmental sequence. Based on the available population data the model can help to estimate appropriate time scales for such measurements.

- The model advocates the view that lineage specification is a continuous process and the hierarchic appearance of hematopoietic differentiation occurs as a consequence of the underlying, self-organizing mechanisms. In this case, it might be insufficient to map the differentiation process onto strictly defined, discrete subpopulations (like short-term repopulating HSCs or common myeloid progenitors that are defined based on a certain expression of surface markers) as the discretization might ignore transient states. However, these transient states might be important intermediates for the understanding of the temporally extended differentiation process.

- The model suggest that reversible developments are rare but at the same time demonstrates that they are an inevitable consequence of the temporally extended and stochastic decision process. Such reversible developments can be typically addressed in experiments with changing culture conditions.

- The model demonstrates that the robust population behavior can be well achieved on the basis of a random and heterogeneous developments on the cellular level. The model indicates that single cell analysis (including a cells progeny over a number of generations) is the most powerful tool to address this population inherent heterogeneity.

- Furthermore the model suggests that the analysis of single cell developments can be used to address functional mechanisms of lineage specification which are not accessible on the population level (e.g. whether certain decision steps are governed by *instructive* versus *selective* mechanisms). In particular, cell death events are important indicators that should not be neglected. The model can also serve as an appropriate reference system.

On the theoretical side, the research on proposed model framework for hematopoietic stem cell organization needs to be continued at different points:

- There are several experimental reports demonstrating that individual HSCs show different and inheritable contributions to lymphoid and myeloid cells types [44, 45] (see also Section 2.3.4). As outlined in Section 5.3 the model of lineage specification is in principle able to reflect aspects of inheritable lineage bias that needs to be adapted to the available data. It is an interesting topic to further analyze how lineage bias and differential repopulation ability of HSCs are functionally coupled to each other.

- The simplified lineage propensities used throughout this thesis might be replaced by small scale interaction networks that appropriately reflect the molecular interaction of key players being involved in cellular decision making.

- The amended stem cell model does not account for any feedback signal from later stages of hematopoietic development (e.g. stimulation of stem cells by cytokines like Erythropoietin or G-CSF). However, for the establishment of a comprehensive blood model both the stem cell compartment as well as the maturing blood and the connections between these different developmental stages need to be represented appropriately.

- A more sophisticated analysis of experimentally derived cellular genealogies should also account for the spatial position and the shape of the tracked cell objects. These features need to be complemented on the modeling side as well.

These extensions of the model and its further applications will help to evaluate the validity of the presented approaches. Advancements as the ones proposed above will further contribute to a better conceptual understanding of tissue stem cell formation on different organizational layers and their mutual interdependence.

Acknowledgement

The presented thesis is based on the research I carried out at the Institute for Medical Informatics, Statistics and Epidemiology at the Medical Faculty of the University of Leipzig, Germany.

I am most thankful to my advisor Dr. Ingo Roeder. He introduced me to the field of theoretical stem cell research and especially in the initial years of my thesis he patiently answered all my questions about stem cells and mathematics. Ingo actively guided my research in many directions, and he was always a highly competent and reliable teacher. With him I had the most fruitful and creative discussions out of which many new ideas were born. I am very thankful that Ingo encouraged and actively supported my future research career by giving me the chance to take own responsibilities and by involving me in many strategic decisions.

In equal measure I am indebted to Prof. Markus Loeffler. His superior knowledge and his long-standing experience provided much support to the work of my thesis. With his enthusiasm and his spirit for new ideas together with his critical and enquiring discussions of our manuscripts, Prof. Loeffler gave a strong impact to my research. As the head of the Institute he furthermore provided all the necessary administrative and financial support for the research of this thesis and related projects. I very much enjoyed the friendly and pleasant atmosphere in his institute.

I also want to express my thanks to the other members of the research group on "Dynamical Modelling of Tissue Stem Cell Organization", in particular to Katrin Horn, Matthias Horn, Ronny Lorenz and Lars Thielecke. Katrin, Matthias and Lars helped a lot with many questions and challenges regarding the simulation code. I am especially thankful to Ronny who completely transferred my crude implementation of the cell tracking analysis into a robust and powerful computational platform. The modularity and the finite running times of the *gAnalyzer* are attributed to his impressive programming skills.

I am furthermore thankful to Dr. Dirk Hasenclever who greatly supported the discussions on the statistical measures for cellular genealogies that are outlined in Chapter 6 and 7.

I grateful acknowledge the support from Dr. Michael Cross who invested the time and took me to his laboratory. Although I could not please him with my experimental skills I personally learnt a lot from these weeks in the "wet" lab, and also from the stimulating discussion we had.

I want to thank my family and my friends for the their support and their motivation during the whole time of my thesis. In particular I want to mention my parents Rosemarie and Rüdiger Glauche, Veronika's parents Traute and Günther Börner as well as my friend Lena Bachmann.

The picture at the front page was especially designed for this thesis by my friend Christian Kroll. I thank him for this beautiful and generous gift.

I am especially thankful to Markus Loeffler and Ingo Roeder for providing the financial support during the time of my PhD thesis and for giving me the chance to attend quite a number of extra-ordinary and stimulating conferences and workshops.

Finally, my most intimate thanks are dedicated to my daughters Elisabeth and Henriette that were both born during the time of my thesis and to their mother and my beloved partner Veronika Börner. It is only their presence that adds a personal value to all this work.

Bibliography

[1] C. S. Potten and M. Loeffler, *Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. lessons for and from the crypt.*, Development **110**, 1001 (1990).

[2] I. Lemischka, in *Handbook of Stem Cells*, edited by R. Lanza, J. Gearhart, B. Hogan, D. Melton, R. Pedersen, J. Thomson, and M. West (Academic Press, Burlington, 2004), pp. 21–46.

[3] M. J. Evans and M. H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.*, Nature **292**, 154 (1981).

[4] G. R. Martin, *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.*, Proc. Natl. Acad. Sci. U. S. A. **78**, 7634 (1981).

[5] J. A. Thomson and V. S. Marshall, *Primate embryonic stem cells.*, Curr Top Dev Biol **38**, 133 (1998).

[6] K. Takahashi and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.*, Cell **126**, 663 (2006).

[7] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.*, Cell **131**, 861 (2007).

[8] K. Okita, T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.*, Nature **448**, 313 (2007).

[9] M. Wernig, A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B. E. Bernstein, and R. Jaenisch, *In vitro reprogramming of fibroblasts into a pluripotent es-cell-like state.*, Nature **448**, 318 (2007).

[10] A. D. Metcalfe and M. W. J. Ferguson, *Tissue engineering of replacement skin: the crossroads of biomaterials, wound healing, embryonic development, stem cells and regeneration.*, J R Soc Interface **4**, 413 (2007).

[11] R. M. Blaese, K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, et al., *T*

lymphocyte-directed gene therapy for ada- scid: initial trial results after 4 years., Science **270**, 475 (1995).

[12] S. Hacein-Bey-Abina, F. L. Deist, F. Carlier, C. Bouneaud, C. Hue, J.-P. D. Villartay, A. J. Thrasher, N. Wulffraat, R. Sorensen, S. Dupuis-Girod, et al., *Sustained correction of x-linked severe combined immunodeficiency by ex vivo gene therapy.*, N Engl J Med **346**, 1185 (2002).

[13] M. G. Ott, M. Schmidt, K. Schwarzwaelder, S. Stein, U. Siler, U. Koehl, H. Glimm, K. Kühlcke, A. Schilz, H. Kunkel, et al., *Correction of x-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of mds1-evi1, prdm16 or setbp1.*, Nat Med **12**, 401 (2006).

[14] S. Hacein-Bey-Abina, C. V. Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, et al., *Lmo2-associated clonal t cell proliferation in two patients after gene therapy for scid-x1.*, Science **302**, 415 (2003).

[15] D. Orlic, J. Kajstura, S. Chimenti, I. Jakoniuk, S. M. Anderson, B. Li, J. Pickel, R. Mckay, B. Nadal-ginard, D. M. Bodine, et al., *Bone marrow cells regenerate infarcted myocardium.*, Nature **410**, 701 (2001).

[16] K. R. Chien, *Stem cells: lost in translation.*, Nature **428**, 607 (2004).

[17] N. D. Theise, S. Badve, R. Saxena, O. Henegariu, S. Sell, J. M. Crawford, and D. S. Krause, *Derivation of hepatocytes from bone marrow cells in mice after radiation-induced myeloablation.*, Hepatology **31**, 235 (2000).

[18] T. R. Brazelton, F. M. Rossi, G. I. Keshet, and H. M. Blau, *From marrow to brain: expression of neuronal phenotypes in adult mice*, Science **290**, 1775 (2000).

[19] E. Mezey, K. J. Chandross, G. Harta, R. A. Maki, and S. R. McKercher, *Turning blood into brain: cells bearing neuronal antigens generated in vivo from bone marrow.*, Science **290**, 1779 (2000).

[20] G. Ferrari, G. C.-D. Angelis, M. Coletta, E. Paolucci, A. Stornaiuolo, G. Cossu, and F. Mavilio, *Muscle regeneration by bone marrow-derived myogenic progenitors.*, Science **279**, 1528 (1998).

[21] M. Dezawa, H. Ishikawa, Y. Itokazu, T. Yoshihara, M. Hoshino, S. Takeda, C. Ide, and Y. Nabeshima, *Bone marrow stromal cells generate muscle cells and repair muscle degeneration.*, Science **309**, 314 (2005).

[22] T. Sato, J. H. Laver, and M. Ogawa, *Reversible expression of cd34 by murine hematopoietic stem cells*, Blood **94**, 2548 (1999).

[23] S. Matsuoka, Y. Ebihara, M. Xu, T. Ishii, D. Sugiyama, H. Yoshino, T. Ueda, A. Manabe, R. Tanaka, Y. Ikeda, et al., *Cd34 expression on long-term repopulating hematopoietic stem cells changes during developmental stages.*, Blood **97**, 419 (2001).

[24] I. Kim, S. He, O. H. Yilmaz, M. J. Kiel, and S. J. Morrison, *Enhanced purification of fetal liver hematopoietic stem cells using slam family receptors.*, Blood **108**, 737 (2006).

[25] C. Eckfeldt, E. Mendenhall, and C. Verfaillie, *The molecular repertoire of the 'almighty' stem cell.*, Nat Rev Mol Cell Biol **6**, 726 (2005).

[26] T. Suda, J. Suda, and M. Ogawa, *Single-cell origin of mouse hemopoietic colonies expressing multiple lineages in variable combinations.*, Proc Natl Acad Sci U S A **80**, 6689 (1983).

[27] T. Suda, J. Suda, and M. Ogawa, *Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors.*, Proc Natl Acad Sci U S A **81**, 2520 (1984).

[28] A. G. Leary, L. C. Strauss, C. I. Civin, and M. Ogawa, *Disparate differentiation in hemopoietic colonies derived from human paired progenitors.*, Blood **66**, 327 (1985).

[29] M. A. Cross and T. Enver, *The lineage commitment of haemopoietic progenitor cells.*, Curr Opin Genet Dev **7**, 609 (1997).

[30] M. Hu, D. Krause, M. Greaves, S. Sharkis, M. Dexter, C. Heyworth, and T. Enver, *Multilineage gene expression precedes commitment in the hemopoietic system.*, Genes Dev **11**, 774 (1997).

[31] T. Enver and M. Greaves, *Loops, lineage, and leukemia.*, Cell **4**, 9 (1998).

[32] K. Akashi, X. He, J. Chen, H. Iwasaki, C. Niu, B. Steenhard, J. Zhang, J. Haug, and L. Li, *Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis.*, Blood **101**, 383 (2003).

[33] I. Roeder and M. Loeffler, *A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity*, Exp. Hematol. **30**, 853 (2002).

[34] M. A. Kirkland, *A phase space model of hemopoiesis and the concept of stem cell renewal.*, Exp Hematol **32**, 511 (2004).

[35] J. Suda, T. Suda, and M. Ogawa, *Analysis of differentiation of mouse hemopoietic stem cells in culture by sequential replating of paired progenitors.*, Blood **64**, 393 (1984).

[36] R. K. Humphries, A. C. Eaves, and C. J. Eaves, *Self-renewal of hemopoietic stem cells during mixed colony formation in vitro.*, Proc Natl Acad Sci U S A **78**, 3629 (1981).

[37] M. Ward, J. Milledge, and J. West, *High Altitude Medicine and Physiology* (Oxford University Press, London, 2000).

[38] D. Douer and H. P. Koeffler, *Retinoic acid enhances growth of human early erythroid progenitor cells in vitro.*, J Clin Invest **69**, 1039 (1982).

[39] C. M. Heyworth, T. M. Dexter, O. Kan, and A. D. Whetton, *The role of hemopoietic growth factors in self-renewal and differentiation of il-3-dependent multipotential stem cells.*, Growth Factors **2**, 197 (1990).

[40] L. Bruno, R. Hoffmann, F. McBlane, J. Brown, R. Gupta, C. Joshi, S. Pearson, T. Seidl, C. Heyworth, and T. Enver, *Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro.*, Mol Cell Biol **24**, 741 (2004).

[41] A. G. Rolink, S. L. Nutt, F. Melchers, and M. Busslinger, *Long-term in vivo reconstitution of t-cell development by pax5-deficient b-cell progenitors.*, Nature **401**, 603 (1999).

[42] C. Heyworth, S. Pearson, G. May, and T. Enver, *Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells.*, EMBO. J. **21**, 3770 (2002).

[43] K. Akashi, *Lineage promiscuity and plasticity in hematopoietic development.*, Ann N Y Acad Sci **1044**, 125 (2005).

[44] C. Muller-Sieburg, R. Cho, L. Karlsson, J. Huang, and H. Sieburg, *Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness.*, Blood **103**, 4111 (2004).

[45] B. Dykstra, D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S.-J. Lee, R. Brinkman, and C. Eaves, *Long-term propagation of distinct hematopoietic differentiation programs in vivo.*, Cell Stem Cell **1**, 218 (2007).

[46] M. Loeffler and I. Roeder, *Tissue stem cells: Definition, plasticity, heterogeneity, self-organization and models - a conceptual approach*, Cells Tissues Organs **171**, 8 (2002).

[47] I. Roeder, *Dynamical modeling of hematopoietic stem cell organization*, Ph.D. thesis, University of Leipzig, Germany (2003).

[48] J. E. Till and E. A. McCulloch, *A direct measurement of radiation sensitivity of normal mouse bone marrow cells*, Radiat. Res. **14**, 213 (1961).

[49] A. J. Becker, E. A. McCulloch, and J. E. Till, *Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells.*, Nature **197**, 452 (1963).

[50] L. Siminovitch, E. A. McCulloch, and J. E. TILL, *The distribution of colony-forming cells among spleen colonies.*, J Cell Physiol **62**, 327 (1963).

[51] M. Ramalho-Santos and H. Willenbring, *On the origin of the term "stem cell".*, Cell Stem Cell **1**, 35 (2007).

[52] L. Alonso and E. Fuchs, *Stem cells of the skin epithelium.*, Proc Natl Acad Sci U S A **100 Suppl 1**, 11830 (2003).

[53] C. S. Potten and R. J. Morris, *Epithelial stem cells in vivo.*, J Cell Sci Suppl **10**, 45 (1988).

[54] N. D. Theise, *Liver stem cells: prospects for treatment of inherited and acquired liver diseases.*, Expert Opin Biol Ther **3**, 403 (2003).

[55] S. Temple, *The development of neural stem cells.*, Nature **414**, 112 (2001).

[56] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones, *Embryonic stem cell lines derived from human blastocysts.*, Science **282**, 1145 (1998).

[57] N. B. Ivanova, J. T. Dimos, C. Schaniel, J. A. Hackney, K. A. Moore, and I. R. Lemischka, *A stem cell molecular signature.*, Science **298**, 601 (2002).

[58] M. Ramalho-Santos, S. Yoon, Y. Matsuzaki, R. C. Mulligan, and D. A. Melton, *"stemness": transcriptional profiling of embryonic and adult stem cells.*, Science **298**, 597 (2002).

[59] H. Mikkers and J. Frisén, *Deconstructing stemness.*, EMBO J **24**, 2715 (2005).

[60] H. M. Blau, T. R. Brazelton, and J. M. Weimann, *The evolving concept of a stem cell: entity or function?*, Cell **105**, 829 (2001).

[61] D. Zipori, *The nature of stem cells: state rather than entity.*, Nat Rev Genet **5**, 873 (2004).

[62] M. Loeffler and C. S. Potten, *Stem cells and cellular pedigrees - a conceptual introduction* (Academic Press, Cambridge, 1997), pp. 1–27.

[63] H. K. A. Mikkola and S. H. Orkin, *The journey of developing hematopoietic stem cells.*, Development **133**, 3733 (2006).

[64] S. H. Orkin, *Diversification of haematopoietic stem cells to specific lineages.*, Nat Rev Genet **1**, 57 (2000).

[65] S. H. Cheshier, S. J. Morrison, X. Liao, and I. L. Weissman, *In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells.*, Proc. Natl. Acad. Sci. U. S. A. **96**, 3120 (1999).

[66] A. Wilson, E. Laurenti, G. Oser, R. C. van der Wath, W. Blanco-Bose, M. Jaworski, S. Offner, C. F. Dunant, L. Eshkind, E. Bockamp, et al., *Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair.*, Cell **135**, 1118 (2008).

[67] I. Glauche, K. Moore, L. Thielecke, K. Horn, M. Loeffler, and I. Roeder, *Stem cell proliferation and quiescence–two sides of the same coin.*, PLoS Comput Biol **5**, e1000447 (2009).

[68] J. L. Abkowitz, S. N. Catlin, M. T. McCallie, and P. Guttorp, *Evidence that the number of hematopoietic stem cells per animal is conserved in mammals.*, Blood **100**, 2665 (2002).

[69] G. de Haan, S. J. Szilvassy, T. E. Meyerrose, B. Dontje, B. Grimes, and G. Van Zant, *Distinct functional properties of highly purified hematopoietic stem cells from mouse strains differing in stem cell numbers.*, Blood **96**, 1374 (2000).

[70] D. C. Weksberg, S. M. Chambers, N. C. Boles, and M. A. Goodell, *Cd150-side population cells represent a functionally distinct population of long-term hematopoietic stem cells.*, Blood **111**, 2444 (2008).

[71] C. L. Celso, H. E. Fleming, J. W. Wu, C. X. Zhao, S. Miake-Lye, J. Fujisaki, D. Cote, D. W. Rowe, C. P. Lin, and D. T. Scadden, *Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche.*, Nature **457**, 92 (2009).

[72] J. Trentin, *Influence of hematopoietic organ stroma (hematopoieticinductive microenvironment) on stem cell differentiation.* (Appleton-Century-Crofts, New York, 1970), pp. 161–168.

[73] J. J. Trentin, *Determination of bone marrow stem cell differentiation by stromal hemopoietic inductive microenvironments (HIM).*, Am. J. Pathol. **65**, 621 (1971).

[74] R. Schofield, *The relationship between the spleen colony-forming cell and the haemopoietic stem cell.*, Blood Cells **4**, 7 (1978).

[75] D. L. Jones and A. J. Wagers, *No place like home: anatomy and function of the stem cell niche.*, Nat Rev Mol Cell Biol **9**, 11 (2008).

[76] M. Cross, R. Alt, and D. Niederwieser, *The case for a metabolic stem cell niche.*, Cells Tissues Organs **188**, 150 (2008).

[77] K. A. Moore and I. R. Lemischka, *Stem cells and their niches.*, Science **311**, 1880 (2006).

[78] F. Arai, A. Hirao, M. Ohmura, H. Sato, S. Matsuoka, K. Takubo, K. Ito, G. Y. Koh, and T. Suda, *Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche.*, Cell **118**, 149 (2004).

[79] A. Czechowicz, D. Kraft, I. L. Weissman, and D. Bhattacharya, *Efficient transplantation via antibody-based clearance of hematopoietic stem cell niches.*, Science **318**, 1296 (2007).

[80] J. L. Christensen, D. E. Wright, A. J. Wagers, and I. L. Weissman, *Circulation and chemotaxis of fetal hematopoietic stem cells.*, PLoS Biol **2**, E75 (2004).

[81] H. Taniguchi, T. Toyoshima, K. Fukao, and H. Nakauchi, *Presence of hematopoietic stem cells in the adult liver.*, Nat Med **2**, 198 (1996).

[82] M. J. Kiel, O. H. Yilmaz, T. Iwashita, O. H. Yilmaz, C. Terhorst, and S. J. Morrison, *Slam family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells.*, Cell **121**, 1109 (2005).

[83] T. Yin and L. Li, *The stem cell niches in bone.*, J Clin Invest **116**, 1195 (2006).

[84] M. Osawa, K. Hanada, H. Hamada, and H. Nakauchi, *Long-term lymphohematopoietic reconstitution by a single cd34- low/negative hematopoietic stem cell.*, Science **273**, 242 (1996).

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

[86] I. L. Weissman, *Stem cells: units of development, units of regeneration, and units in evolution.*, Cell **100**, 157 (2000).

[87] A. J. Wagers, J. L. Christensen, and I. L. Weissman, *Cell fate determination from stem cells.*, Gene Ther **9**, 606 (2002).

[88] D. Metcalf, *Concise review: hematopoietic stem cells and tissue stem cells: current concepts and unanswered questions.*, Stem Cells **25**, 2390 (2007).

[89] H. E. Broxmeyer and D. E. Williams, *The production of myeloid blood cells and their regulation during health and disease.*, Crit Rev Oncol Hematol **8**, 173 (1988).

[90] H. E. Broxmeyer and C. H. Kim, *Regulation of hematopoiesis in a sea of chemokine family members with a plethora of redundant activities.*, Exp Hematol **27**, 1113 (1999).

[91] M. Haggstrom, picture published under GNU Free Documentation License at http://en.wikipedia.org/wiki/File:Hematopoiesis_simple.png.

[92] S. Huang, Y.-P. Guo, G. May, and T. Enver, *Bifurcation dynamics in lineage-commitment in bipotent progenitor cells.*, Dev Biol **305**, 695 (2007).

[93] C. Joshi and T. Enver, *Molecular complexities of stem cells.*, Curr. Opin. Hematol. **10**, 220 (2003).

[94] S. J. Morrison, N. M. Shah, and D. J. Anderson, *Regulatory mechanisms in stem cell biology.*, Cell **88**, 287 (1997).

[95] T. Enver, C. M. Heyworth, and T. M. Dexter, *Do stem cells play dice?*, Blood **92**, 348 (1998).

[96] T. Graf, *Differentiation plasticity of hematopoietic cells.*, Blood **99**, 3089 (2002).

[97] C. M. Heyworth, M. Alauldin, M. A. Cross, L. J. Fairbairn, T. M. Dexter, and A. D. Whetton, *Erythroid development of the FDCP-Mix A4 multipotent cell line is governed by the relative concentrations of erythropoietin and interleukin 3.*, Br J Haematol **91**, 15 (1995).

[98] N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I. R. Lemischka, *Dissecting self-renewal in stem cells with rna interference.*, Nature **442**, 533 (2006).

[99] A. B. Cantor and S. H. Orkin, *Transcriptional regulation of erythropoiesis: an affair involving multiple partners.*, Oncogene **21**, 3368 (2002).

[100] M. A. Cross, C. M. Heyworth, A. M. Murrell, E. O. Bockamp, T. M. Dexter, and A. R. Green, *Expression of lineage restricted transcription factors precedes lineage specific differentiation in a multipotent haemopoietic progenitor cell line.*, Oncogene **9**, 3013 (1994).

[101] M. H. Sieweke and T. Graf, *A transcription factor party during blood cell differentiation.*, Curr Opin Genet Dev **8**, 545 (1998).

[102] J. E. Visvader, M. Crossley, J. Hill, S. H. Orkin, and J. M. Adams, *The c-terminal zinc finger of gata-1 or gata-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line.*, Mol Cell Biol **15**, 634 (1995).

[103] C. Nerlov and T. Graf, *Pu.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors.*, Genes Dev **12**, 2403 (1998).

[104] C. Nerlov, E. Querfurth, H. Kulessa, and T. Graf, *GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription.*, Blood **95**, 2543 (2000).

[105] N. Rekhtman, F. Radparvar, T. Evans, and A. I. Skoultchi, *Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells.*, Genes Dev **13**, 1398 (1999).

[106] N. Rekhtman, K. S. Choe, I. Matushansky, S. Murray, T. Stopka, and A. I. Skoultchi, *PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation.*, Mol Cell Biol **23**, 7460 (2003).

[107] P. Zhang, G. Behre, J. Pan, A. Iwama, N. Wara-Aswapati, H. S. Radomska, P. E. Auron, D. G. Tenen, and Z. Sun, *Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1.*, Proc Natl Acad Sci U S A **96**, 8705 (1999).

[108] I. Roeder and I. Glauche, *Towards an understanding of lineage specification in hematopoietic stem cells: A mathematical model for the interaction of transcription factors GATA-1 and PU.1.*, J Theor Biol **241**, 852 (2006).

[109] P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, and H. Singh, *Multilineage transcriptional priming and determination of alternate hematopoietic cell fates.*, Cell **126**, 755 (2006).

[110] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Construction of a genetic toggle switch in Escherichia coli.*, Nature **403**, 339 (2000).

[111] O. Cinquin and J. Demongeot, *Positive and negative feedback: striking a balance between necessary antagonists.*, J Theor Biol. **216**, 229 (2002).

[112] N. D. Theise and R. Harris, *Stem Cells* (Springer-Verlag Berlin Heidelberg, 2006), chap. Postmodern Biology: (Adult) (Stem) Cells Are Plastic, Stochastic, Complex, and Uncertain, pp. 389–408.

[113] D. Zipori, *The stem state: plasticity is essential, whereas self-renewal and hierarchy are optional.*, Stem Cells **23**, 719 (2005).

[114] I. Glauche, M. Cross, M. Loeffler, and I. Roeder, *Lineage specification of hematopoietic stem cells: Mathematical modeling and biological implications.*, Stem Cells **25**, 1791 (2007).

[115] H. Xie, M. Ye, R. Feng, and T. Graf, *Stepwise reprogramming of b cells into macrophages.*, Cell **117**, 663 (2004).

[116] L. Chen, H. Zhang, Y. Shi, K. L. Chin, D. C. Tang, and G. P. Rodgers, *Identification of key genes responsible for cytokine-induced erythroid and myeloid differentiation and switching of hematopoietic stem cells by rage.*, Cell Res **16**, 923 (2006).

[117] C. Brawley and E. Matunis, *Regeneration of male germline stem cells by spermatogonial dedifferentiation in vivo.*, Science **304**, 1331 (2004).

[118] T. Kai and A. Spradling, *Differentiating germ cells can revert into functional stem cells in drosophila melanogaster ovaries.*, Nature **428**, 564 (2004).

[119] S. Knaan-Shanzer, I. van der Velde-van Dijke, M. J. M. van de Watering, P. J. de Leeuw, D. Valerio, D. W. van Bekkum, and A. A. F. de Vries, *Phenotypic and functional reversal within the early human hematopoietic compartment.*, Stem Cells **26**, 3210 (2008).

[120] H. B. Sieburg, R. H. Cho, B. Dykstra, N. Uchida, C. J. Eaves, and C. E. Muller-Sieburg, *The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets.*, Blood **107**, 2311 (2006).

[121] C. E. Muller-Sieburg and H. B. Sieburg, *Clonal diversity of the stem cell compartment.*, Curr Opin Hematol **13**, 243 (2006).

[122] I. Roeder, K. Horn, H.-B. Sieburg, R. Cho, C. Muller-Sieburg, and M. Loeffler, *Characterization and quantification of clonal heterogeneity among hematopoietic stem cells: a model-based approach.*, Blood **112**, 4874 (2008).

[123] H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang, *Transcriptome-wide noise controls lineage choice in mammalian progenitor cells.*, Nature **453**, 544 (2008).

[124] T. Schroeder, *Tracking hematopoiesis at the single cell level.*, Ann N Y Acad Sci **1044**, 201 (2005).

[125] I. Glauche, R. Lorenz, D. Hasenclever, and I. Roeder, *A novel view on stem cell development: analysing the shape of cellular genealogies.*, Cell Prolif **42**, 248 (2009).

[126] B. Dykstra, J. Ramunas, D. Kent, L. McCaffrey, E. Szumsky, L. Kelly, K. Farn, A. Blaylock, C. Eaves, and E. Jervis, *High-resolution video monitoring of hematopoietic stem cells cultured in single-cell arrays identifies new features of self-renewal.*, Proc Natl Acad Sci U S A **103**, 8185 (2006).

[127] M. Punzel, D. Liu, T. Zhang, V. Eckstein, K. Miesala, and A. D. Ho, *The symmetry of initial divisions of human hematopoietic progenitors is altered only by the cellular microenvironment.*, Exp. Hematol. **31**, 339 (2003).

[128] O. Al-Kofahi, R. J. Radke, S. K. Goderie, Q. Shen, S. Temple, and B. Roysam, *Automated cell lineage construction: a rapid method to analyze clonal development established with murine neural progenitor cells.*, Cell Cycle **5**, 327 (2006).

[129] B. M. Deasy, R. J. Jankowski, T. R. Payne, B. Cao, J. P. Goff, J. S. Greenberger, and J. Huard, *Modeling stem cell population growth: incorporating terms for proliferative heterogeneity.*, Stem Cells **21**, 536 (2003).

[130] J. Ramunas, H. J. Montgomery, L. Kelly, T. Sukonnik, J. Ellis, and E. J. Jervis, *Real-time fluorescence tracking of dynamic transgene variegation in stem cells.*, Mol Ther **15**, 810 (2007).

[131] M. Stadtfeld and T. Graf, *Assessing the role of hematopoietic plasticity for endothelial and hepatocyte development by non-invasive lineage tracing.*, Development **132**, 203 (2005).

[132] J. Zhang, F. Varas, M. Stadtfeld, S. Heck, N. Faust, and T. Graf, *Cd41-yfp mice allow in vivo labeling of megakaryocytic cells and reveal a subset of platelets hyperreactive to thrombin stimulation.*, Exp Hematol **35**, 490 (2007).

[133] B. I. Lord, *Biology of the haemopoietic stem cell* (Academic Press, Cambridge, 1997), pp. 401–422.

[134] R. H. Cho and C. E. Muller-Sieburg, *High frequency of long-term culture-initiating cells retain in vivo repopulation and self-renewal capacity.*, Exp. Hematol. **28**, 1080 (2000).

[135] R. P. van Os, B. Dethmers-Ausema, and G. de Haan, *In vitro assays for cobblestone area-forming cells, ltc-ic, and cfu-c.*, Methods Mol Biol **430**, 143 (2008).

[136] N. Uchida, W. H. Fleming, E. J. Alpern, and I. L. Weissman, *Heterogeneity of hematopoietic stem cells.*, Curr. Opin. Immunol. **5**, 177 (1993).

[137] M. A. Goodell, K. Brose, G. Paradis, A. S. Conner, and R. C. Mulligan, *Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo.*, J Exp Med **183**, 1797 (1996).

[138] L. Gallacher, B. Murdoch, D. M. Wu, F. N. Karanu, M. Keeney, and M. Bhatia, *Isolation and characterization of human CD34(-)Lin(-) and CD34(+)Lin(-) hematopoietic stem cells using cell surface markers AC133 and CD7.*, Blood **95**, 2813 (2000).

[139] M. A. Goodell, *Cd34+ or cd34-: does it really matter?*, Blood **94**, 2545 (1999).

[140] S. Soneji, S. Huang, M. Loose, I. J. Donaldson, R. Patient, B. Göttgens, T. Enver, and G. May, *Inference, validation, and dynamic modeling of transcription networks in multipotent hematopoietic cells.*, Ann N Y Acad Sci **1106**, 30 (2007).

[141] V. Munugalavadla and R. Kapur, *Role of c-kit and erythropoietin receptor in erythropoiesis.*, Crit Rev Oncol Hematol **54**, 63 (2005).

[142] M. O. Arcasoy and X. Jiang, *Co-operative signalling mechanisms required for erythroid precursor expansion in response to erythropoietin and stem cell factor.*, Br J Haematol **130**, 121 (2005).

[143] A. Rosmarin, Z. Yang, and K. Resendes, *Transcriptional regulation in myelopoiesis: Hematopoietic fate choice, myeloid differentiation, and leukemogenesis.*, Exp Hematol **33**, 131 (2005).

[144] N. Drize, J. Chertkov, E. Sadovnikova, S. Tiessen, and A. Zander, *Long-term maintenance of hematopoiesis in irradiated mice by retrovirally transduced peripheral blood stem cells.*, Blood **89**, 1811 (1997).

[145] O. S. Kustikova, C. Baum, and B. Fehse, *Retroviral integration site analysis in hematopoietic stem cells.*, Methods Mol Biol **430**, 255 (2008).

[146] A. Gerrits, B. Dykstra, O. J. Kalmykowa, K. Klauke, E. Verovskaya, M. J. C. Broekhuis, G. de Haan, and L. V. Bystrykh, *Cellular barcoding tool for clonal analysis in the hematopoietic system.*, Blood, Epub ahead of print (2010).

[147] T. Schroeder, *Imaging stem-cell-driven regeneration in mammals.*, Nature **453**, 345 (2008).

[148] K. H. Campbell, J. McWhir, W. A. Ritchie, and I. Wilmut, *Sheep cloned by nuclear transfer from a cultured cell line.*, Nature **380**, 64 (1996).

[149] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, et al., *Induced pluripotent stem cell lines derived from human somatic cells.*, Science **318**, 1917 (2007).

[150] H. Zhou, S. Wu, J. Y. Joo, S. Zhu, D. W. Han, T. Lin, S. Trauger, G. Bien, S. Yao, Y. Zhu, et al., *Generation of induced pluripotent stem cells using recombinant proteins.*, Cell Stem Cell **4**, 381 (2009).

[151] H. Stachowiak, *Allgemeine Modelltheorie [in german]* (Springer, Wien, 1973).

[152] M. Loeffler and I. Roeder, *Conceptual models to understand tissue stem cell organization.*, Curr Opin Hematol **11**, 81 (2004).

[153] R. Gross and M. Loeffler, *Prinzipien der Medizin. [in german]* (Springer, Berlin, 1997).

[154] S. Viswanathan and P. W. Zandstra, *Towards predictive models of stem cell fate.*, Cytotechnology **41**, 75 (2003).

[155] J. E. Till, E. A. McCulloch, and L. Siminovitch, *A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells*, Proc. Natl. Acad. Sci. **51**, 29 (1964).

[156] M. C. Mackey and L. Glass, *Oscillation and chaos in physiological control systems*, Science **197**, 287 (1977).

[157] M. C. Mackey, *Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis.*, Blood **51**, 941 (1978).

[158] M. Loeffler and H. E. Wichmann, *A comprehensive mathematical model of stem cell proliferation which reproduces most of the published experimental results.*, Cell Tissue. Kinet. **13**, 543 (1980).

[159] U. Reincke, M. Loeffler, H. E. Wichmann, and B. Harrison, *The kinetics of granulopoiesis in long-term mouse bone marrow culture. part i.*, Int J Cell Cloning **2**, 394 (1984).

[160] H. E. Wichmann, M. Loeffler, and U. Reincke, *The kinetics of granulopoiesis in long-term mouse bone marrow culture. part ii.*, Int J Cell Cloning **2**, 408 (1984).

[161] M. Loeffler, K. Pantel, H. Wulff, and H. E. Wichmann, *A mathematical model of erythropoiesis in mice and rats. part 1: Structure of the model.*, Cell Tissue Kinet **22**, 13 (1989).

[162] H. E. Wichmann, M. Loeffler, K. Pantel, and H. Wulff, *A mathematical model of erythropoiesis in mice and rats. part 2: Stimulated erythropoiesis.*, Cell Tissue Kinet **22**, 31 (1989).

[163] H. Wulff, H. E. Wichmann, K. Pantel, and M. Loeffler, *A mathematical model of erythropoiesis in mice and rats. part 3: Suppressed erythropoiesis.*, Cell Tissue Kinet **22**, 51 (1989).

[164] M. Loeffler, R. Stein, H. E. Wichmann, C. S. Potten, P. Kaur, and S. Chwalinski, *Intestinal cell proliferation. i. a comprehensive model of steady-state proliferation in the crypt.*, Cell Tissue Kinet **19**, 627 (1986).

[165] M. Loeffler, C. S. Potten, and H. E. Wichmann, *Epidermal cell proliferation. ii. a comprehensive mathematical model of cell proliferation and migration in the basal layer predicts some unusual properties of epidermal stem cells.*, Virchows Arch B Cell Pathol Incl Mol Pathol **53**, 286 (1987).

[166] M. Loeffler and B. Grossmann, *A stochastic branching model with formation of subunits applied to the growth of intestinal crypts.*, J. Theor. Biol. **150**, 175 (1991).

[167] M. Ogawa and T. R. Mosmann, in *Leukemia: Recent Advances in Biology and Treatment* (Alan R. Liss Inc., 1985), pp. 391–397.

[168] P. Guttorp, M. A. Newton, and J. L. Abkowitz, *A stochastic model for haematopoiesis in cats.*, IMA. J. Math. Appl. Med. Biol. **7**, 125 (1990).

[169] J. L. Abkowitz, S. N. Catlin, and P. Guttorp, *Evidence that hematopoiesis may be a stochastic process in vivo.*, Nat. Med. **2**, 190 (1996).

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

[171] S. N. Catlin, P. Guttorp, and J. L. Abkowitz, *The kinetics of clonal dominance in myeloproliferative disorders.*, Blood **106**, 2688 (2005).

[172] G. Colvin, J. Lambert, M. Abedi, M. Dooner, D. Demers, B. Moore, D. Greer, J. Aliotta, J. Pimentel, J. Cerny, et al., *Differentiation hotspots: the deterioration of hierarchy and stochasm.*, Blood Cells Mol Dis **32**, 34 (2004).

[173] G. A. Challen, N. Boles, K. K.-Y. Lin, and M. A. Goodell, *Mouse hematopoietic stem cell identification and analysis.*, Cytometry A **75**, 14 (2009).

[174] N. Theise and M. d'Inverno, *Understanding cell lineages as complex adaptive systems.*, Blood Cells Mol Dis **32**, 17 (2004).

[175] P. J. Quesenberry, G. A. Colvin, and J. F. Lambert, *The chiaroscuro stem cell: a unified stem cell theory.*, Blood **100**, 4266 (2002).

[176] A. Kurakin, *Self-organization vs watchmaker: stochastic gene expression and cell differentiation.*, Dev Genes Evol **215**, 46 (2005).

[177] H. Vogel, H. Niewisch, and G. Matioli, *Stochastic development of stem cells.*, J. Theor. Biol. **22**, 249 (1969).

[178] T. Hearn, C. Haurie, and M. C. Mackey, *Cyclical neutropenia and the peripheral control of white blood cell production.*, J. Theor. Biol. **192**, 167 (1998).

[179] C. Colijn and M. C. Mackey, *A mathematical model of hematopoiesis–i. periodic chronic myelogenous leukemia.*, J Theor Biol **237**, 117 (2005).

[180] C. Colijn and M. C. Mackey, *A mathematical model of hematopoiesis: Ii. cyclical neutropenia.*, J Theor Biol **237**, 133 (2005).

[181] M. Scholz, C. Engel, and M. Loeffler, *Modelling human granulopoiesis under poly-chemotherapy with g-csf support.*, J Math Biol **50**, 397 (2005).

[182] L. J. Fairbairn, G. J. Cowling, B. M. Reipert, and T. M. Dexter, *Suppression of apoptosis allows differentiation and development of a multipotent hemopoietic cell line in the absence of added growth factors.*, Cell **74**, 823 (1993).

[183] I. Roeder and R. Lorenz, *Asymmetry of stem cell fate and the potential impact of the niche observations, simulations, and interpretations.*, Stem Cell Rev **2**, 171 (2006).

[184] J. Betschinger and J. A. Knoblich, *Dare to be different: asymmetric cell division in drosophila, c. elegans and vertebrates.*, Curr Biol **14**, R674 (2004).

[185] M. Goetz and W. B. Huttner, *The cell biology of neurogenesis.*, Nat Rev Mol Cell Biol **6**, 777 (2005).

[186] T. Lechler and E. Fuchs, *Asymmetric cell divisions promote stratification and differentiation of mammalian skin.*, Nature **437**, 275 (2005).

[187] M. Hoffmann, H. H. Chang, S. Huang, D. E. Ingber, M. Loeffler, and J. Galle, *Noise-driven stem cell and progenitor population dynamics.*, PLoS One **3**, e2922 (2008).

[188] C. H. Waddington, *Principles of Embryology* (George Allen & Unwin, Ltd., 1956).

[189] C. H. Waddington, *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology* (George Allen & Unwin, Ltd., 1957).

[190] D. E. Ingber, *Tensegrity ii. how structural networks influence cellular information processing networks.*, J Cell Sci **116**, 1397 (2003).

[191] R. K. Ng, W. Dean, C. Dawson, D. Lucifero, Z. Madeja, W. Reik, and M. Hemberger, *Epigenetic restriction of embryonic cell lineage fate by methylation of elf5.*, Nat Cell Biol **10**, 1280 (2008).

[192] B. Goldmann, *Smash the (cell) state!*, Nature Reports Stem Cells, doi:10.1038/stemcells.2008.115 (2008).

[193] S. Yamanaka, *Elite and stochastic models for induced pluripotent stem cell generation.*, Nature **460**, 49 (2009).

[194] P. Ao, *Global view of bionetwork dynamics: adaptive landscape.*, J Genet Genomics **36**, 63 (2009).

[195] Y. Cao and J. Liang, *Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability.*, BMC Syst Biol **2**, 30 (2008).

[196] H. Preisler and S. Kauffman, *A proposal regarding the mechanism which underlies lineage choice during hematopoietic differentiation.*, Leuk Res **23**, 685 (1999).

[197] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection of Evolution.* (Oxford University Press, 1993).

[198] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber, *Cell fates as high-dimensional attractor states of a complex gene regulatory network.*, Phys Rev Lett **94**, 128701 (2005).

[199] A. Wuensche, *Modularity in Development and Evolution* (Chicago University Press, 2004), chap. Basins of Attraction in Network Dynamics: A Conceptual Framework for Biomolecular Networks, pp. 288–311.

[200] S. Kauffman, *Homeostasis and differentiation in random genetic control networks.*, Nature **224**, 177 (1969).

[201] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic gene expression in a single cell.*, Science **297**, 1183 (2002).

[202] M. Kaern, T. Elston, W. Blake, and J. Collins, *Stochasticity in gene expression: from theories to phenotypes.*, Nat Rev Genet **6**, 451 (2005).

[203] J. M. Raser and E. K. O'Shea, *Noise in gene expression: origins, consequences, and control.*, Science **309**, 2010 (2005).

[204] C. Metzner, M. Sajitz-Hermstein, M. Schmidberger, and B. Fabry, *Noise and critical phenomena in biochemical signaling cycles at small molecule numbers*, Phys Rev E, (accepted) (2009).

[205] S. Huang, *Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery.*, J Mol Med **77**, 469 (1999).

[206] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, *Noise-based switches and amplifiers for gene expression.*, Proc Natl Acad Sci U S A **97**, 2075 (2000).

[207] T. Tian and K. Burrage, *Stochastic models for regulatory networks of the genetic toggle switch.*, Proc Natl Acad Sci U S A **103**, 8372 (2006).

[208] G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz, *Tunability and noise dependence in differentiation dynamics.*, Science **315**, 1716 (2007).

[209] R. Losick and C. Desplan, *Stochasticity and cell fate.*, Science **320**, 65 (2008).

[210] M. Ogawa, *Differentiation and proliferation of hematopoietic stem cells.*, Blood **81**, 2844 (1993).

[211] K. Kaneko and T. Yomo, *Isologous diversification: a theory of cell differentiation.*, Bull Math Biol **59**, 139 (1997).

[212] C. Furusawa and K. Kaneko, *Origin of complexity in multicellular organisms.*, Phys Rev Lett **84**, 6130 (2000).

[213] C. Furusawa and K. Kaneko, *Theory of robustness of irreversible differentiation in a stem cell system: chaos hypothesis.*, J Theor Biol **209**, 395 (2001).

[214] H. Takagi and K. Kaneko, *Dynamical systems basis of metamorphosis: diversity and plasticity of cellular states in reaction diffusion network.*, J Theor Biol **234**, 173 (2005).

[215] D. G. Tenen, *Disruption of differentiation in human cancer: AML shows the way.*, Nat Rev Cancer **3**, 89 (2003).

[216] L. Mariani, M. Loehning, A. Radbruch, and T. Hoefer, *Transcriptional control networks of cell differentiation: insights from helper t lymphocytes.*, Prog Biophys Mol Biol **86**, 45 (2004).

[217] A. Yates, R. Callard, and J. Stark, *Combining cytokine signalling with t-bet and gata-3 regulation in th1 and th2 differentiation: a model for cellular decision-making.*, J Theor Biol **231**, 181 (2004).

[218] O. Cinquin and J. Demongeot, *High-dimensional switches and the modelling of cellular differentiation.*, J Theor Biol **233**, 391 (2005).

[219] D. V. Foster, J. G. Foster, S. Huang, and S. A. Kauffman, *A model of sequential branching in hierarchical cell fate determination.*, J Theor Biol **260**, 589 (2009).

[220] I. Roeder, M. Horn, I. Glauche, A. Hochhaus, M. C. Mueller, and M. Loeffler, *Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications.*, Nat Med **12**, 1181 (2006).

[221] I. Roeder, L. Kamminga, K. Braesel, B. Dontje, G. de Haan, and M. Loeffler, *Competitive clonal hematopoiesis in mouse chimeras explained by a stochastic model of stem cell organization.*, Blood **105**, 609 (2005).

[222] H. Takano, H. Ema, K. Sudo, and H. Nakauchi, *Asymmetric division and lineage commitment at the level of hematopoietic stem cells: inference from differentiation in daughter cell and granddaughter cell pairs.*, J Exp Med **199**, 295 (2004).

[223] C. Luecke, *Quantitative modellierung des stammzellverhaltens unter in-vitro-bedingungen am beispiel von einzelzellinduzierten cafc-assays [in german]*, M.D. thesis, University of Leipzig, Germany (2008).

[224] C. Engel, M. Scholz, and M. Loeffler, *A computational model of human granulopoiesis to simulate the hematotoxic effects of multicycle polychemotherapy.*, Blood **104**, 2323 (2004).

[225] H. M. Eilken, S.-I. Nishikawa, and T. Schroeder, *Continuous single-cell imaging of blood generation from haemogenic endothelium.*, Nature **457**, 896 (2009).

[226] M. A. Rieger, P. S. Hoppe, B. M. Smejkal, A. C. Eitelhuber, and T. Schroeder, *Hematopoietic cytokines can instruct lineage choice.*, Science **325**, 217 (2009).

[227] A. Mooers and S. Heard, *Inferring evolutionary process from the phylogenetic tree shape.*, Q Rev Biol **72**, 31 (1997).

[228] M. Kirkpatrick and M. Slatkin, *Searching for evolutionary patterns in the shape of a phylogenetic tree.*, Evolution **47**, 1171 (1993).

[229] P.-M. Agapow and A. Purvis, *Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis.*, Syst Biol **51**, 866 (2002).

[230] D. H. Colless, *Phylogenetics: the theory and practice of phylogenetic systematics ii.*, Syst. Zool **31**, 100 (1982).

[231] M. A. Rieger and T. Schroeder, *Exploring hematopoiesis at single cell resolution.*, Cells Tissues Organs **188**, 139 (2008).

[232] N. Scherf, I. Roeder, and I. Glauche, in *Proceedings of the Fifth International Workshop on Computational Systems Biology, WCSB 2008, Leipzig, Germany* (2008), pp. 161 – 164.

[233] D. Hume, *Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression.*, Blood **96**, 2323 (2000).

[234] V. Chickarmane, T. Enver, and C. Peterson, *Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility.*, PLoS Comput Biol **5**, e1000268 (2009).

[235] A. Krinner, M. Zscharnack, A. Bader, D. Drasdo, and J. Galle, *Impact of oxygen environment on mesenchymal stem cell expansion and chondrogenic differentiation.*, Cell Prolif **42**, 471 (2009).

[236] Z. Agur, Y. Daniel, and Y. Ginosar, *The universal properties of stem cells as pinpointed by a simple discrete model.*, J. Math. Biol. **44**, 79 (2002).

[237] W. A. Prudhomme, K. H. Duggar, and D. A. Lauffenburger, *Cell population dynamics model for deconvolution of murine embryonic stem cell self-renewal and differentiation responses to cytokines and extracellular matrix.*, Biotechnol Bioeng **88**, 264 (2004).

Appendix A.

Details of the update procedure

A.1. The *Pólya* urn model

The general idea of the outlined update procedure for the lineage propensities, i.e. the random choice of one lineage with a probability equal to the relative abundance of its propensity and the subsequent update of this lineage, is primarily inspired by the so called *Pólya urn model*. The original problem was stated by the Hungarian mathematician George Pólya for an urn initially containing a number of black and white marbles. After one marble is chosen randomly from the urn, it is put back into the urn together with another marble of the same color (corresponding to the stated reward m_i). As no normalization procedure is applied, the total number of marbles in the urn grows continuously and in the course of time one color outcompetes the other (i.e. one type of marbles is contained in much higher frequency as compared to the other). This scenario is closely resembled in the dissipative control regime of the proposed lineage specification model.

A.2. Renormalization of the vector of lineage propensities

The normalization constant C^n defined in Equation (4.7) can be expressed as

$$C^n = 1 + m_i = \frac{1}{1-\lambda} \tag{A.1}$$

Under the condition, that m_i for the updated lineage is small ($m_i \approx 0$), the value of the newly introduced constant λ is small as well.

Using this constant in Equation (4.6) the update step reads:

$$\begin{aligned} \mathbf{x}(t+1) &= \frac{1}{C^n}\mathbf{x}(t)(1+\bar{m}) \\ &= \mathbf{x}(t)(1+\bar{m})(1-\lambda) \\ &= \mathbf{x}(t)(1+\bar{m}-\lambda+\lambda\mathbf{x}) \\ &\approx \mathbf{x}(t)(1+\bar{m}-\lambda) \end{aligned} \tag{A.2}$$

In this formulation it is evident that the normalization corresponds to a first order decay term acting on all lineage propensities. However, the interpretation

of the decay rate λ requires some care. First, the decay rate is not a constant as λ is a function of $m_i = f_m(x_i(t))$ which is itself a function of the actual propensities $\mathbf{x}(t)$. Thus λ has to be recalculated in every update step.

Second, as the lineage specific reward m_i can be negative in the *regressive control regime* the decay rate λ is negative as well. In this case, the lineage i chosen for the update procedure is finally decreased whereas the other competing lineages increase their lineage propensities. For this scenario the notion of a decay rate fails.

Appendix B.

Modifications of the original HSC model

Prior and in parallel to the implementation of the model of lineage specification a number of modifications have been made to the numerical implementation of the model of HSC self-renewal and differentiation as proposed by Roeder [47]. A brief overview of the changes is provided below.

- **Universal list of all cells.** Within the original model structure by Roeder [47] dormant cells in $\mathcal{A}$, activated cells in $\Omega\,(a > a_{\min})$ and differentiating cells (termed clones, $a < a_{\min}$) were treated in separate lists. However, as the cell's attributes are the same for all these cells they are now represented in a universal list termed `totalCell`. For each cell element within this list the function `simulation()` in `sc_model.cpp` calls the update function termed `updateCell()`. This update function, defined in `sc_cell_class.cpp`, distinguishes whether the cell is in $\mathcal{A}$ or Ω and calls the corresponding update function `updateDormant()` or `updateActive()`, respectively.

- **Proliferating and maturating cells.** Cells with $a < a_{\min}$ are treated as normal cells in the universal list of cells. However, these cells undergo a period of further divisions with turnover time `clone_cycle_time` and duration `proliferation_span`. The relevant parameters are specified in the `input` file. The proliferation period is followed by a maturation period, in which the cells do not undergo further divisions. After a total time (`clone_life_span`, measured in time steps (typically hours) after the cell passed $a = a_{\min}$) the cell is finally removed from the list of cells.

  For the convenient description of these time ranges a linear scale parallel to the logarithmic scale in terms of the affinity parameter a has been introduced. The novel scale, termed `ageOfLine`, is constructed such that `ageOfLine`$= 0$ for $a = a_{\min}$. `ageOfLine` is measured in time units of the update process (typically hours). E.g. for a cell with $a = 1$ the corresponding negative value for `ageOfLine` corresponds to the number of time steps the cell needs to reduce their affinity until $a = a_{\min}$, given a certain differentiation rate d.

- **Multiple cells.** For numerical efficiency, a cell object in the list of all cells might not represent just a single cell but a group of ν cells, indicated by the counter `myActualCellNumber` of the `cell` class. Technically,

upon division the cells undergo a virtual division by multiplying the counter `myActualCellNumber` by a factor 2 as compared to the generation of a new cell object in the list of all cells.

Application of this routine requires the adaption of further functions for the simulation of binding (change from Ω to $\mathcal{A}$) and cell death.

- **Cell counting.** For keeping track of the actual cell numbers in each of the simulated compartments a general function, termed `countCell()`, has been put into place. This function is called after each update step of a cell and upon generation of new cells during division.

- **Unlimited number of cell lines.** The program code has been extended to simulate a (in principle unlimited) number of different cell types with different dynamical characteristics. The specifications have to be made in the `input` file by providing the number of different cell types (`number_of_types`) and a sufficient parameter set for each cell type.

- **Cell tracking.** For the analysis of cellular genealogies the tracks of each individual cell need to be extracted from the simulation model. The corresponding functionality is provided by the class `sc_historytree_class.cpp`. Upon specification in the `sequence` file the tracking information of individual cells is extracted and stored in a corresponding output file. For each cell and each time step a complete record of the cell's defining parameters (i.e. growth environment, attachment affinity a, lineage propensities $\mathbf{x}$) is stored along with the information of the divisional history (i.e. information about the parent cell and the daughter cells).

- **Boundary at $a_{\min}$.** As the original distinction between cells in Ω on whether they correspond to the list of activated cells ($a > a_{\min}$) or to the list of differentiating clones ($a < a_{\min}$) has been abandoned; the relevant, common update function `updateActive()` has been reformulated. Upon evaluation of a cell's probability to change into $\mathcal{A}$ the cell's affinity a is multiplied with a step function f_s. This step function approaches $f_s(a > a_{\min}) \to 1$ and $f_s(a < a_{\min}) \to 0$ in the opposite case. The steepness of the step function is defined in the `input` file by the parameter `steepness`.

The programming of these changes and the establishment of the new simulation package called *SCMuni* has been performed with great support by Katrin Horn, Matthias Horn, Ronny Lorenz and Lars Thielecke.

Appendix C.

Computational aspects of the model of lineage specification

C.1. Extension of the SCMuni code

The relevant parameters for the simulation of lineage specification are defined in the `input` files. The general simulation of lineage specification is under governance of the flag parameter `lineageSpecification`.

The following parameter correspondence has been applied:

parameter definition in the `input` file	**corresponding parameter throughout the thesis**
`number_of_lines`	N
`initDiffState`	initial propensities $\mathbf{x}(t=0)$
`ResponceActive`	reward in the dissipative control regime $m_i = n_i$
`ResponceDormant`	saturation parameter of the sigmoid reward function in the regressive control regime σ_i
`rewardDormantSlope`	steepness parameter of the sigmoid/linear reward function in the regressive control regime b_i
`lowerResponceXaxis`	root of the sigmoid/linear reward function in the regressive control regime x^R
`selectionMean`	lower bound of the cell death process $x_i^{\text{death/low}}$
`selectionVar`	`selectionMean` + `selectionVar` add up to the upper bound of the cell death process $x_i^{\text{death/high}}$
`selectionAmplitude`	lineage specific cell death intensity ϕ_i
`selectionBackground`	background cell death Φ^B affecting all cells

The central function of the lineage specification within the *SCMuni* code is the function `lineageDevelopment()`. The function belongs to the `Cell` class and is generally called within the `updateCell()` function. Within this and a number of secondary functions the actual vector representing the lineage propensities (named `myDiffState`) is updated according to the general model outlined in Section 4.3 of the main text.

The *selective* mode of lineage specification and the background cell death are implemented by use of the functions `lineageSelection()` and `backgroundApoptosis()`, respectively. Both functions belong to the `Cell` class and are called within the `updateCell()` function, too.

The general output of the lineage specification process is written to a file named `scm_*_DiffCountSummary.dat`. This file contains a list in which for every time step the total number of uncommitted cells ($x_i^*(t) < x_{\mathrm{com1}}$) and cells committed to lineage i ($x_i^*(t) > x_{\mathrm{com1}}$) is provided. Further output (`scm_*_StatAminTransition.dat`, `scm_*_StatFinalCommit.dat`, `scm_*_PoolDifferentialCount.dat`) can be generated by invocing the function `enhancedStatistics` in the input file. Descriptions of these specifically tailored output files is contained in their headers.

C.2. Analysis of the cellular genealogies

As briefly outlined in Appendix B the tracks of cellular history are written to a particular file specified in the `sequence` file. As these tracking files contain information about every tracked cell at each time point their file size increases exponentially with the simulation time.

For the vizualization and the analysis of the resulting cellular genealogies a specific tool, called *gAnalzyer*, has been developed and implemented in *C++*.

The *gAnalyzer* provides the following functionalities:

- processing of the raw tracking files provided by the *SCMuni* code
- internal representation of the cellular genealogies in a tree like structure (each cell element has a reference to its parental cell and, possibly, to its daughter cells, as well as a container with all relevant tracking information such as cell specific parameters as a function of time; this way, the topology is formally separated from the additional information)
- analysis of cellular genealogies using the methods introduced in Chapter 6 with text-based output
- framework for the development of further, possible recursive analysis methods for cellular genealogies
- graphical representation of the cellular genealogies using different coloring schemes and file formats (such as jpeg, gif, eps)

The *gAnalyzer* has been implemented, tested and documented by Ronny Lorenz. The finite running times and the modularity of this valuable tool are attributed to his programming skills.

C.3. Statistical and graphical analysis of the results

Graphics are generated using the software packages *gnuplot* and *R*. The package *R* was also used for statistical analysis in Chapter 7.

Sketches are produced using the programs *xfig* and *inkscape*.

The *SCMuni* as well as the *gAnalyzer* a programmed in *C++*. A number of additional scripts to control sequential simulation runs are written in *Perl*.

Appendix D.

Simulation protocols and parameters

D.1. Studies on the regressive control regime

For the simulation results on the analysis of the regressive control regime shown in Section 5.1.1, a truncated version of the Roeder and Loeffler model is used. In this version, cells are artificially kept in signaling context $\mathcal{A}$. This is achieved by setting the transition probability to signaling context Ω to 0. Thus, these cells remain under the continuous governance of the regressive control regime. Further parameters of the Roeder and Loeffler model, as they are discussed in Section 3.4 are of no relevance for the particular simulations.

Parameters for the regressive reward function are provided in the corresponding Figure legends and the main text in Section 5.1.1.

D.2. Studies on the dissipative control regime

For the simulation results on the analysis of the dissipative control regime shown in Section 5.1.2, a similar, truncated version of the Roeder and Loeffler model is applied as in Section 5.1.1. However, studying the development in the dissipative control regime, cells are artificially kept in signaling context Ω. This is achieved by setting the transition probability to signaling context $\mathcal{A}$ to 0 and by reducing the differentiation coefficient $d \rightarrow 1$. Furthermore, cell division is disabled and cells remain in G1 for the time of study. Further parameters of the Roeder and Loeffler model are of no relevance for the particular simulations.

Parameters for the dissipative reward function are provided in the Figure legends and the main text in Section 5.1.2.

D.3. Studies on the extended model system

For the simulation results shown for the extended model in Section 5.2 and all following model simulations a standard parameter set for the Roeder and Loeffler model is used which has been adapted for the simulation of B6 mice [221]. The particular parameters are provided in Table D.1. Differences from the standard values as well as information on the parameters governing lineage specification are provided in the Figure legends and the main text.

D.4. Simulation results: Lineage contribution of single differentiating cells

Simulation results on lineage contribution of single cells *in vitro* are provided in Section 5.4.1.

Model parameters for the *source assays* are given in Supplementary Table D.2. These assays are initialized with 250 cells in signaling context A and 100 cells in signaling context Ω such that the system reaches a homeostatic situation within the initial simulation time t_{sim}. For further transfer, cells within the transfer range a_{trans} from N_{pool} realizations of the source assay are pooled together (*transfer pool*). This way the influence of system inherent fluctuations is minimized. The transfer pool S and T, defined by the transfer range a^T_{trans}, are provided in Supplementary Table D.2.

Parameters of the *lineage assay* are given in Supplementary Table D.3. N_{transfer} cells from the *transfer pool* are chosen randomly and continue cell cycle at position $c = c_{\text{reenter}}$. Committed cells ($x_i^* > x_{\text{com}} = 0.9$) are excluded from the transfer pool. After expansion for t_{sim2} hours the composition of the progeny of the daughter cells is evaluated. Cells with $x_i^* > x_{\text{com}} = 0.9$ are counted as finally committed cells of the particular lineage i and add to the lineage contribution of the particular clone.

The critical parameter of the dissipative control regime $n_i^{\text{prog}} = m_i^{\text{prog}} = 0.15$ was chosen such that the leading lineage propensity x_i^* of the large majority of cells reaches the level $x_{\text{com}} = 0.9$ within four to ten days corresponding to the occurrence of lineage specific markers or characteristic morphological changes. Given the low sensitivity of the results on the choice of parameters in the regressive control regime, parameter values are used which allow moderate deviations of the lineage propensities from the mean level to converge within a period of one to two days.

For the actual fit of the model results, the size of the transfer ranges $a^{S/T}_{\text{trans}}$ has been adapted to meet the experimental data of the discussed critical experiments by Suda et al. [26, 27] and Takano et al. [222]. This corresponds to the modification of the composition of cells in the transfer pools S and T from which cells are chosen for the individual tracking. Besides the reduction from $N = 6$ (Suda et al. [27]) to $N = 4$ (Takano et al. [222]) simulated lineages, all other parameters are left unchanged.

For the simulation of developmental correlations between certain lineages, introduced in Figure 5.20, a correlation coefficient $\gamma_{14} = \gamma_{41} = 0.3$ is applied in both the source and the lineage assay. Details are provided in the main text.

D.5. Simulation results: Comparative differentiation of paired daughter cells

Simulation results on lineage contribution of paired daughter cells *in vitro* are provided in Section 5.4.2.

Model parameters for the particular *source assays* are identical to those in the case of the single cell differentiation given in Supplementary Table D.2. Additionally, the same transfer pools S and T are applied for the case of the paired daughter experiments. This is consistent with the experimental situation in which the same selection protocols are used for both types of experiments (compare [26, 27] and [222]).

Parameters of the simulated *division assays* are provided in Supplementary Table D.4. N_{parent} single cells from the transfer pool are used for transfer into the division assays. Committed cells ($x_i^* > x_{\text{com}} = 0.9$) are excluded from the transfer pool. Transplanted cells in the division assay are always placed under the governance of the proliferative signaling context Ω, all other individual parameters are maintained. Within the division assay transition to signaling context Ω is impaired for numerical simplicity. Daughter cells are cultured for $t_{\text{transplant}}$ hours after division until they are further transferred into the lineage assays.

Simulation parameters of the final *lineage assay* are identical to those in the case of single cell differentiation provided in Supplementary Table D.3. As in the previous set of simulation, cells with $x_i^* > x_{\text{com}} = 0.9$ are counted as finally committed cells of the particular lineage i and add to the lineage contribution of the particular clone.

D.6. Simulation results: Lineage specification in differentiating FDCP-mix cells

The simulation for the lineage specification in differentiating FDCP-mix cell cultures are presented in Section 5.4.3. In particular, the simulations are performed using both the *instructive* and the *selective* mode of lineage specification. Parameters of the simulated *differentiation assays* are provided in Tables D.5 and D.6

For the *instructive* mode, cells with $x_i^* > x_{\text{com2}} = 0.9$ are counted as finally committed cells of the particular lineage i except for the intermediate stage of erythroblasts for which this level is decrease to $x_{\text{com1/E}} = 0.7$. Adapting the model to the *selective* scenario, these values are slightly decreased. Thus, cells with $x_i^* > x_{\text{com2}} = 0.85$ are counted as finally committed cells and for the intermediate stage of erythroblasts this level is decrease to $x_{\text{com1/E}} = 0.6$.

The lineage specific rewards of the dissipative control regime $m_i^{\text{prog}} = n_i^{\text{prog}}$ in the *instructive* mode and the lineage specific cell death amplitudes in the *selective*

mode are used to fit the simulation results individually to the data on M and E differentiation. The initial pool of cells a_{init} has been adapted such that no self-renewing cells are present after three days of culture. This is suggested by particular CFU assays carried out in parallel to the differentiation assays.

D.7. Simulation results: Scenarios for the generation of cellular genealogies

A detailed description of the simulation routines and used parameters is provided for the three general classes of simulation scenarios presented in Section 7.1.

Growth Scenario. For the growth scenario 400 model systems are initialized, each with an individual cell. These individual cells are tracked for 300 hours, and subsequently the cellular genealogies are derived. Model parameters are chosen such that the system can establish a homeostatic situation. The parameters are given in Supplementary Table D.7.

Homeostatic scenario. Initializing a model system with a single cell, cell numbers reach a homeostatic situation after about 600 hours (compare Figure 3 in the main publication). At time point t=700 hours, all stem cells (i.e. cells with $a > 0.1$) of a particular realization are uniquely marked and subsequently tracked for the next 300 hours. For the particular, randomly chosen realization 399 cells with affinity $a > 0.1$ have been found and their genealogies have been reconstructed. Model parameters are identical to the growth scenario. The differences in the genealogies result from the differences in the initial configuration (expansion of a single cell in an empty model system versus homeostasis in a "filled" system). Parameters are given in Supplementary Table D.7.

Differentiation scenario. For the differentiation scenario, all stem cells (i.e. cells with $a > 0.1$) of a particular realization are uniquely marked in the homeostatic situation at time point t=1500. In the next step the differentiation rate is increased to $d = 1.1$ and the regeneration rate is reduced to $r = 1.0$ for all cells. This way, the cells rapidly decrease their affinity a, lose the potential for self-renewal and undergo terminal differentiation. 390 cells with $a > 0.1$ have been found in the particular realization. These cells have been tracked during the next 300 h, and genealogies have been reconstructed. Parameters of the simulations are given in Supplementary Table D.7.

D.8. Simulation results: Instructive versus selective lineage specification

For the simulations in Section 7.2, 500 cells are initialized in $\mathcal{A}$ with affinity $a = 0.5$ and lineage propensities $x_{1,2} = 0.5$. In the *instructive* scenario, overall commitment towards lineage $i = 2$ is achieved by tuning the lineage specific rewards n_i in the dissipative control regime. In contrast, a similar commitment in the *selective* scenario is achieved by adjustung the lineage specific, selective cell death rates ϕ_i. Parameters are adjusted such that the overall dynamics of the lineage specification process appear very similar (compare Figure 7.10). A detailed list of parameters is provided in Supplementary Table D.8.

parameter	value
d	1.07
r	1.1 (1.0 in the *in vitro* situation)
$a_{\min}$	0.01
τ_c	24 hours
τ_S	8 hours
$\tau_{G_2/M}$	4 hours
t_{prolif}	500 hours
t_{mature}	350 hours
$f_A(0)$	0.5
$f_A(\frac{\tilde{N}}{2})$	0.3
$f_A(\tilde{N})$	0.01
$f_A(\infty)$	0
N_A	400 (1300)
$f_\Omega(0)$	0.5
$f_\Omega(\frac{\tilde{N}}{2})$	0.3
$f_\Omega(\tilde{N})$	0.1
$f_\Omega(\infty)$	0
N_Ω	80 (280)

Table D.1.: General model parameters for the simulation of HSC organization in B6 mice.
d: differentiation coefficient; r: regeneration coefficient; τ_c: cell cycle duration; τ_S: S-phase duration; $\tau_{G_2/M}$: duration of G2 and M-phase; t_{prolif}: duration of proliferation phase; t_{mature}: duration of maturation phase; f_A and f_Ω: sigmoid transition functions (see Section 3.4); N_A and N_Ω: norm cell numbers in $\mathcal{A}$ and Ω, respectively; numbers in parenthesis correspond to a enlarged model system with closely similar behavior used for the simulations in Chapter 7. Parameters for the lineage specification model are provided either in the main text and figure legends or following tables D.2 to D.8.

parameter	source assay according to results by Suda et al.[26, 27]	source assay according to results by Takano et al.[222]
d	1.07	1.07
r	1.1	1.1
$a_{\min}$	0.01	0.01
τ_c	24 hours	24 hours
τ_S	8 hours	8 hours
$\tau_{G_2/M}$	4 hours	4 hours
t_{prolif}	250 hours	250 hours
t_{mature}	200 hours	200 hours
N	6	4
dissipative control		
n_i $(i = 1...N)$	0.15	0.15
regressive control		
σ_i $(i = 1...N)$	0.1	0.1
b_i $(i = 1...N)$	-1	-1
x^R	1/6	1/4
t_{sim}	744 hours	744 hours
a_{trans}	$[0.000001, 0.99]$	$[0.012, 0.99]$
N_{pool}	10	50

Table D.2.: Model parameters for the source assays (Sections 5.4.1 and 5.4.2).
N: number of lineage; n_i: lineage specific reward in the dissipative control regime; σ_i, b_i and x^R: parameters of the sigmoid reward function defined in equation (4.4); t_{sim}: length of initial simulations; a_{trans}: affinity range from which cells are chosen for transplantation into further assays; N_{pool}: number of realizations of the source assay used for pooling.

parameter	lineage assay according to results by Suda et al.[27]	lineage assay according to results by Takano et al.[222]
d	1.07	1.07
r	1.0	1.0
$a_{\min}$	0.01	0.01
τ_c	16 hours	16 hours
τ_S	7 hours	7 hours
$\tau_{G_2/M}$	3 hours	3 hours
t_{prolif}	250 hours	250 hours
t_{mature}	200 hours	200 hours
N	6	4
dissipative control		
$n_i\ (i = 1...N)$	0.15	0.15
regressive control		
$\sigma_i\ (i = 1...N)$	0.1	0.1
$b_i\ (i = 1...N)$	-1	-1
x^R	1/6	1/4
c_{reenter}	6 hours	6 hours
t_{sim2}	240 hours	240 hours
N_{transfer}	50,000	50,000

Table D.3.: Model parameters for the lineage assays (Sections 5.4.1 and 5.4.2).

c_{reenter} point in cell cycle, at which cells enter into the lineage assay; t_{sim2} time of evaluation of the lineage assay ("time of culture"). Based on the observation that colonies with more than 2000 cells are formed within 7 days of culture (as reported by Suda et al.[27]), a reduction of the cell cycle time τ_c in the lineage assay is assumed (from 24 to 16 h).

parameter	division assay according to results by Suda et al.[27]	division assay according to results by Takano et al.[222]
d	1.07	1.07
r	1.0	1.0
$a_{\min}$	0.01	0.01
τ_c	24 hours	24 hours
τ_S	8 hours	8 hours
$\tau_{G_2/M}$	4 hours	4 hours
t_{prolif}	250 hours	250 hours
t_{mature}	200 hours	200 hours
N	6	4
dissipative control		
n_i $(i = 1...N)$	0.15	0.15
regressive control		
σ_i $(i = 1...N)$	0.1	0.1
b_i $(i = 1...N)$	-1	-1
x^R	1/6	1/4
N_{parent}	50,000	50,000
$t_{\text{transplant}}$	3 hours	3 hours

Table D.4.: Model parameters for the division assays (Section 5.4.2).
N_{parent} total number of parental cells used for the paired daughter cell simulations; $t_{\text{transplant}}$ time after division at which daughter cells are separated.

parameter	differentiation assay for FDCP-mix cells in M-medium	differentiation assay for FDCP-mix cells in E-medium
d	1.07	1.07
r	1.0	1.0
$a_{\min}$	0.01	0.01
τ_c	24 hours	24 hours
τ_S	8 hours	8 hours
$\tau_{G_2/M}$	4 hours	4 hours
t_{prolif}	150 hours	150 hours
t_{mature}	70 hours	70 hours
N	3	3
dissipative control		
n_1 (G)	0.19	0.078
n_2 (M)	0.155	0.075
n_3 (E)	0.08	0.111
regressive control		
$\sigma_i\,(i = 1...3)$	0.1	0.1
$b_i\,(i = 1...3)$	-1	-1
x^R	1/3	1/3
background cell death		
Φ^B	0	0
N_{init}	250	250
a_{init}	$[0.01, 0.1]$	$[0.01, 0.1]$
t_{sim}	220 hours	220 hours

Table D.5.: Model parameters for FDCP-mix cell differentiation in the instructive mode (Section 5.4.3).
The n_i are given for the particular lineages: G - granulocyte, M - macrophage, E - erythroid. N_{init} refers to the number of initial cells, a_{init} gives the range in which the cells are uniformly seeded.

parameter	differentiation assay for FDCP-mix cells in M-medium	differentiation assay for FDCP-mix cells in E-medium
d	1.07	1.07
r	1.0	1.0
$a_{\min}$	0.01	0.01
τ_c	24 hours	24 hours
τ_S	8 hours	8 hours
$\tau_{G_2/M}$	4 hours	4 hours
t_{prolif}	150 hours	150 hours
t_{mature}	70 hours	70 hours
N	3	3
dissipative control		
$n_i\,(i = 1...3)$	0.11	0.17
regressive control		
$\sigma_i\,(i = 1...3)$	0.1	0.1
$b_i\,(i = 1...3)$	-1	-1
x^R	1/3	1/3
selective cell death		
$x_i^{\text{death/low}}\,(i = 1...3)$	0.5	0.5
$x_i^{\text{death/high}}\,(i = 1...3)$	1	1
ϕ_1 (G)	0.19	0
ϕ_2 (M)	0.155	0.022
ϕ_3 (E)	0.08	0.2
N_{init}	250	250
a_{init}	$[0.01, 0.1]$	$[0.01, 0.1]$
t_{sim}	220 hours	220 hours

Table D.6.: Model parameters for FDCP-mix cell differentiation in the selective mode (Section 5.4.3).
The individual probabilities ϕ_i for the selective cell death process Φ^S differ for the three lineages: $i = 1$ - granulocyte (G), $i = 2$ - macrophage (M), $i = 3$ - erythroid (E).

parameter	growth scenario	homeostatic scenario	differentiation scenario
d	1.07	1.07	1.1
r	1.035	1.035	1.0
$a_{\min}$	0.1	0.1	0.1
τ_c	24 hours	24 hours	24 hours
τ_S	8 hours	8 hours	24 hours
$\tau_{G_2/M}$	4 hours	4 hours	24 hours
t_{prolif}	220 hours	220 hours	220 hours
t_{mature}	175 hours	175 hours	175 hours
N	3	3	3
dissipative control			
n_i $(i = 1...3)$	0.14	0.14	0.14
regressive control			
σ_i $(i = 1...3)$	0.2	0.2	0.2
b_i $(i = 1...3)$	-1	-1	-1
x^R	1/3	1/3	1/3
t_{sim}	300 hours	300 hours	300 hours
tracked cell pool	400 independent simulations each initialized with one cell	one simulation, tracking all 399 cells with $a > a_{\min}$ from homeostatic situation at time point 700	one simulation, tracking all 390 cells with $a > a_{\min}$ from homeostatic situation at time point 1500

Table D.7.: Model parameters for the comparison of cellular genealogies (Section 7.1). t_{sim}: observation period for the individual trajectories.

parameter	instructive	selective
d	1.07	1.07
r	1.0	1.0
$a_{\min}$	0.01	0.01
N_{A}	400	400
N_{Ω}	800	800
τ_c	24 hours	24 hours
τ_S	8 hours	8 hours
$\tau_{G_2/M}$	4 hours	4 hours
t_{prolif}	200 hours	200 hours
t_{mature}	200 hours	200 hours
N	2	2
dissipative control		
n_1	0.043	0.08
n_2	0.065	0.08
regressive control		
$\sigma_{1,2}$	0.1	0.1
$b_{1,2}$	-1	-1
x^R	1/2	1/2
selective cell death		
$x_{1,2}^{\mathrm{death/low}}$	0.6	0.6
$x_{1,2}^{\mathrm{death/high}}$	1	1
ϕ_1	0.0	0.07
ϕ_2	0.0	0
background cell death		
Φ^B	0.0075	0.0
t_{sim}	200 hours	200 hours
initial cell pool	500 cells in Ω with affinity $a = 0.1$ and lineage propensities $x_{1,2} = 0.5$	500 cells in Ω with affinity $a = 0.1$ and lineage propensities $x_{1,2} = 0.5$

Table D.8.: Cellular genealogies under instructive and selective lineage specification (Section 7.2).
Φ^B: background cell death as defined in equation (4.8).

Appendix E.

Selected mathematical abbreviations

$S = \{S_1, S_2, \ldots, S_n\}$.	state space vector
a, c	cell's affinity to signaling context $\mathcal{A}$, position in the cell cycle
$\mathcal{A}$, Ω	signaling contexts
d, r	differentiation, regeneration coefficient
τ_c, τ_{G1}, τ_S, $\tau_{G2/M}$	duration of cell cycle, G1, S, G2/M-phase
α, ω	transition intensities between $\mathcal{A}$ and Ω
f_A, f_Ω	sigmoid transition functions
N_A, N_Ω	norm cell numbers in $\mathcal{A}$ and Ω
N	number of lineages
$\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$.	vector of lineage propensities
$\bar{x} = 1/N$	mean lineage propensity level
x_i^*	lineage propensity of the dominant lineage
m_i	lineage specific reward
$f_m(x_i(t))$	functions defining the lineage specific reward
n_i	constant, lineage specific reward in the dissipative control regime
σ_i, b_i, x^R	parameters of the reward function in the regressive control regime
Φ^S,Φ^B	selective cell death intensity, background cell death intensity
ϕ_i	lineage specific cell death intensities
$x_i^{\text{death/low}}$, $x_i^{\text{death/high}}$.	parameters of the selective cell death intensity
$x_{\text{com1/2}}$	threshold values for the phenotypic mapping

$\mathcal{C} = \{c_i,\, i = 0 \dots n\}$.	set of edges of a cellular genealogy (cells)
$\mathcal{C}^{\text{div}}$	subset of all cells which undergo division
$\mathcal{C}^{\text{death}}$	subset of all cells which undergo cell death
$\mathcal{C}^{\text{term}}$	subset of all cells with censored observation
$\mathcal{C}^{\text{leaf}}$	subset of leaf cells
L	total number of leaf cells in a cellular genealogy
$\mathcal{D} = \{d_j,\, j = 1 \dots m\}$	set of branching points of a cellular genealogy (divisions)
D	total number of divisions in a cellular genealogy
g_i	generation of cell c_i
r_{pq}	topological distance between two cells c_p and c_q
χ_i	fate information for cell c_i
B, B^{char}	branch length, characteristic branch length
B^{range}	range of branch lengths
C, C^w	Colless' index, weighted Colless' index
A, A^g	cell death index, cell death index in generation g
MI	mutual information
R	average minimal distance between characteristic events