

CORRELATION PATTERNS OF CELLULAR GENEALOGIES

Nico Scherf, Ingo Roeder, Ingmar Glauche

Institute for Medical Informatics, Statistics and Epidemiology
University of Leipzig
Haertelstr. 16/18, D-04107 Leipzig, Germany
nico.scherf@ imise.uni-leipzig.de | ingo.roeder@ imise.uni-leipzig.de | ingmar.glauche@ imise.uni-leipzig.de

ABSTRACT

Time lapse video microscopy facilitates the observation and analysis of individual cell fates. Information on cellular development, divisional history, and differentiation are naturally comprised into a pedigree-like structure, denoted as cellular genealogy. Characteristics of the differentiation process are potentially imprinted in these cellular genealogies. Here we study a set of topological measures that are specifically tailored to extract typical correlation patterns between characteristic cell death events and to relate them to relevant biological processes. Using a single-cell based, mathematical model of hematopoietic stem cell organization we compare differentiation strategies that are based on either the instructive or the selective action of cell fate specific signals and show their consequences on the level of the cellular genealogies.

1. INTRODUCTION

Although somatic stem cells play a central role in tissue maintenance and repair as well as in cancer initiation and progression, many questions about their organizational principles are still unresolved. For example, it is an open question whether asymmetric cell division events play a functional role for the maintenance of the stem cell pool or if the observed developmental patterns are induced by asymmetric cell fates which are not necessarily linked to the cell division event [1, 2]. Moreover, the nature of multipotency as well as of the dynamic processes that initiate and regulate the specification of the diversity of functional cells (lineage specification) is only insufficiently understood [3, 4]. In particular, there are reports that lineage specification is an *instructive* process in which a combination of cytokines and cell fate specific signals influences the gene expression pattern of an undifferentiated cell such that certain lineages are promoted whereas others are not. In contrast, it has been argued that lineage specification is *selective* in the sense that the intrinsic influence on the gene expression pattern 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 cell determined towards other lineages are not supported and consequently undergo cell death [1, 3, 5].

Experimental approaches based on cell population averages are mostly not able 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 application of time lapse video microscopy for the analysis of cell cultures facilitates the tracing of single cells, including all its progeny over extended time periods up to several days. This comprises the temporal analysis of cell specific parameters like morphology, cell cycle time, motility or the occurrence of cell death within the population context. All the 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 are referred to as cellular genealogies.

Although the numerical methods are still under development, the automatised analysis of time lapse videos from cell cultures will soon allow the simultaneous tracking of a multitude of root cells. The expected 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. The main objective of this work is the application of a set of recently proposed topological measures [9] to characterize the differences in the cellular genealogies that have been derived using either the *instructive* or the *selective* mode of lineage specification. Since difficulties in the automatic identification and tracing of single cells in current image-processing techniques still limit the availability of experimentally derived cellular genealogies, we use simulated *in silico* cell cul-

tures in order to approach the stated question. In particular, we obtain cellular genealogies from a single-cell based computer-model of hematopoietic stem cell organization which is able to describe self-renewal, differentiation and lineage specification within heterogeneous cell populations and which has been verified for different in vivo and in vitro situations [6, 7, 8]. Based on this model we show how changes in the particular mode of lineage specification (*instructive* vs. *selective*) influence the topology of the cellular genealogies.

2. METHODS

Characterization of cellular genealogies. Cellular genealogies are derived from the tracking of a single, specified cell object (root cell) and its entire clonal offspring.

Technically, a cellular genealogy is an unordered tree graph $\mathcal{G} = \{\mathcal{C}, \mathcal{D}\}$ in which the edges $\mathcal{C} = \{c_i; i = 1, \dots, N\}$ represent cells and the branching points $\mathcal{D} = \{d_i; i = 1, \dots, m\}$ represent 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 \mathcal{G} is uniquely identified by its root cell $c_0 \in \mathcal{C}^0$ which is the cell that had been chosen as the initial cell of the tracking process. Within such a structure cells are ordered into subset \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 g th generation ($c_i \in \mathcal{C}^1, \mathcal{C}^2, \dots$). To each cell c_i belongs either a subsequent division event d_j , giving rise to two daughter cells ($c_i \in \mathcal{C}^{\text{div}}$, with \mathcal{C}^{div} representing the subset of all cells which undergo division), or the cell's existence terminates without a further division either by cell death ($c_i \in \mathcal{C}^{\text{death}}$, with $\mathcal{C}^{\text{death}}$ representing the subset of all cells which die within the observation period) or by termination of the tracking process ($c_i \in \mathcal{C}^{\text{term}}$, with $\mathcal{C}^{\text{term}}$ representing the subset of all cells with censored observation, i.e. no information about future cell fate available). Final cells are termed leaf cells, i.e. $\mathcal{C}^{\text{leaf}} = \mathcal{C}^{\text{death}} \cup \mathcal{C}^{\text{term}}$. The degree of 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 cells c_p and c_q . 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 1.

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 topological parameter.

Generation of cellular genealogies. Cellular genealogies are generated from a single-cell based, mathematical model of hematopoietic stem cell organization that has been developed in our group [6, 7, 8]. Within the model stem cells are able to reversibly switch between two characteristic states: proliferating and quiescent. Cells that have lost their propensity to change into the quiescent state continue regular cell divisions within a proliferation

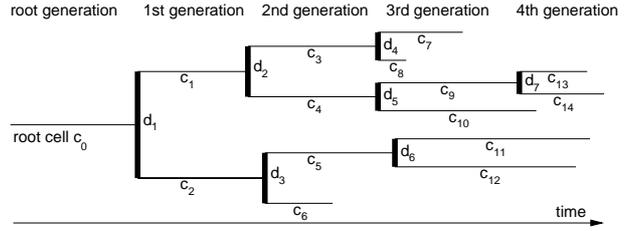


Figure 1. **Schematic sketch of a cellular genealogy.** Within the given five generation genealogy \mathcal{G} the thin horizontal lines represent the cells $c_i \in \mathcal{C}$ whereas the divisions $d_i \in \mathcal{D}$ are marked by the thick vertical bars. The horizontal dimension is time t with the founding root cell c_0 indicated on the left side. 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).

phase (differentiating cells) and are finally removed from the system after a subsequent maturation phase without further divisions.

In this model lineage specification is described by intracellular propensities for the development of particular lineage fates. Whereas the quiescent state equalizes the lineage specific propensities (uncommitted state), the dominance of one or another lineage is established in a stochastic process during proliferation, indicating the process of lineage commitment. In particular it has been assumed that bi-potent progenitor cells are influenced by the (in silico) conditions such that only one of the two possible lineage fates is promoted and the other one is largely suppressed. For the scope of this work two different modes of lineage specification have been applied. In the *selective* mode, we assume that the cell-intrinsic commitment process is unbiased and promotes the development of both possible lineages. However, there is a targeted cell death process preferentially affecting cells that initiated development towards the suppressed lineage, whereas the preferred lineage is largely unaffected. In contrast, in the *instructive* mode, the cell-intrinsic commitment process is biased towards the preferred lineage. In this scenario, cell death occurs randomly in all cells. The parameter configuration has been chosen such that the population kinetics are indistinguishable for both scenarios (Figure 2).

For the application of a number of statistical measures we compare two sets of 500 cellular genealogies, derived either under the *instructive* or the *selective* mode of lineage specification. In particular, we have initiated two cell populations of 500 initially undifferentiated, bi-potent cells with impaired self-renewal ability which undergo the desired lineage specification process generating one of the two possible cell types. The tracking process for each of the genealogies extends over 200 hours.

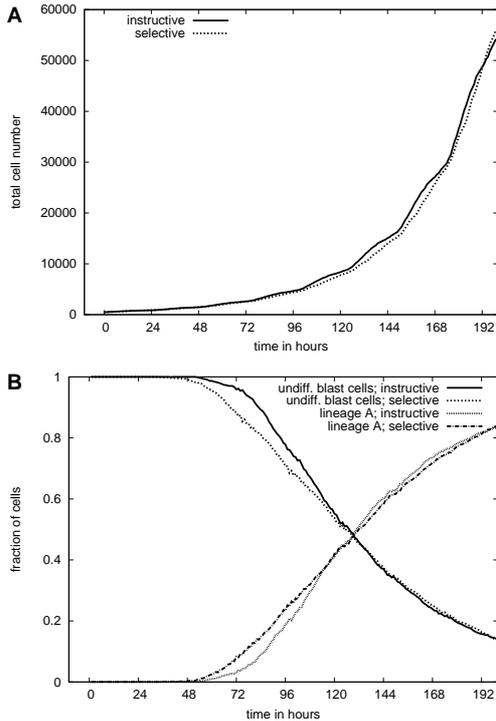


Figure 2. **Population development.** The growth kinetics (A) and the temporal development of the lineage specification (B) are shown for both cell populations (500 initial cells in either the *instructive* or *selective* mode of lineage specification). In (B) the decline of undifferentiated blast cells and the appearance of committed cells of generic type A are provided.

3. RESULTS

Addressing the structural differences in the shapes of the genealogies generated by use of either the *instructive* or the *selective* mode of lineage specification we apply a set of measures that we described previously [9]. In particular the measures on the total number of leaves and the characteristic path lengths are indicators of the expansion. In Figure 3A a boxplot for the distribution of the number of leaves ($L = |\mathcal{C}^{\text{leaf}}|$) is provided for both modes of lineage specification. Since the population kinetics in Figure 2 have been fitted to resemble almost identical growth behavior of the cell cultures, these findings are reflected on the single cell basis, too. Also for the application of weighted Colless' index C^w in Figure 3B, which is a normalized measure of the imbalance within the tree branches, no significant differences in the frequency of occurrence for the 500 sample genealogies can be found.

However, as we have outlined previously, it is the proximity between cell death events which can potentially reveal whether two events are correlated or not. In particular, one assumes, that closely related cells share a similar stage of development, such that these cells undergo similar regulating processes, like induced cell death due to selective lineage specification. In this scenario, cell death

events should occur closer to each other, and more often, preferentially in sibling cells.

In Figure 3C we show a boxplot for the distribution of the distance between a cell death event and the closest other cell death event ($r_p = \min_q(r_{pq}; c_p, c_q \in \mathcal{C}^{\text{death}})$), averaged over all “dead cells” within a particular genealogy. It is obvious that the *selective* mode of lineage specification leads to shorter average minimal distances between such cell death events. Furthermore, we have analyzed the 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). As the corresponding boxplots in Figure 3D indicate, this fraction is increased for the *selective* mode of lineage specification as compared to the *instructive* mode.

4. DISCUSSION

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. We showed that cellular genealogies bear a number of additional information which is not accessible on the population level. We demonstrated that the application of suitable measures, such as the average minimal distance between cell death events or the fraction death sibling cells, is appropriate to distinguish different modes of lineage specification. In particular, the *selective* mode of lineage specification is characterized by an increased fraction of death siblings as compared to the *instructive* mode while the average minimal distance between cell death events is considerably reduced.

We are aware that the application of the outlined measures to a set of experimentally derived cellular genealogies does not ultimately allow the identification of the particular mode of lineage specification since the necessary reference scenario is missing. However, we take this as a strong argument in favor of our modeling approach. Given the population kinetics for the cell culture in question, the mathematical model can be adapted using either the *instructive* or the *selective* mode of lineage specification. The resulting genealogies can act as the reference scenarios to which the experimental data is finally compared.

5. ACKNOWLEDGEMENT

The authors thank Dirk Hasenclever and Ronny Lorenz for discussion and programming. This research was funded by European Commission project EuroSyStem (200270).

6. REFERENCES

- [1] T. Schroeder, “Tracking hematopoiesis at the single cell level.,” *Ann N Y Acad Sci*, vol. 1044, pp. 201–9, 2005.

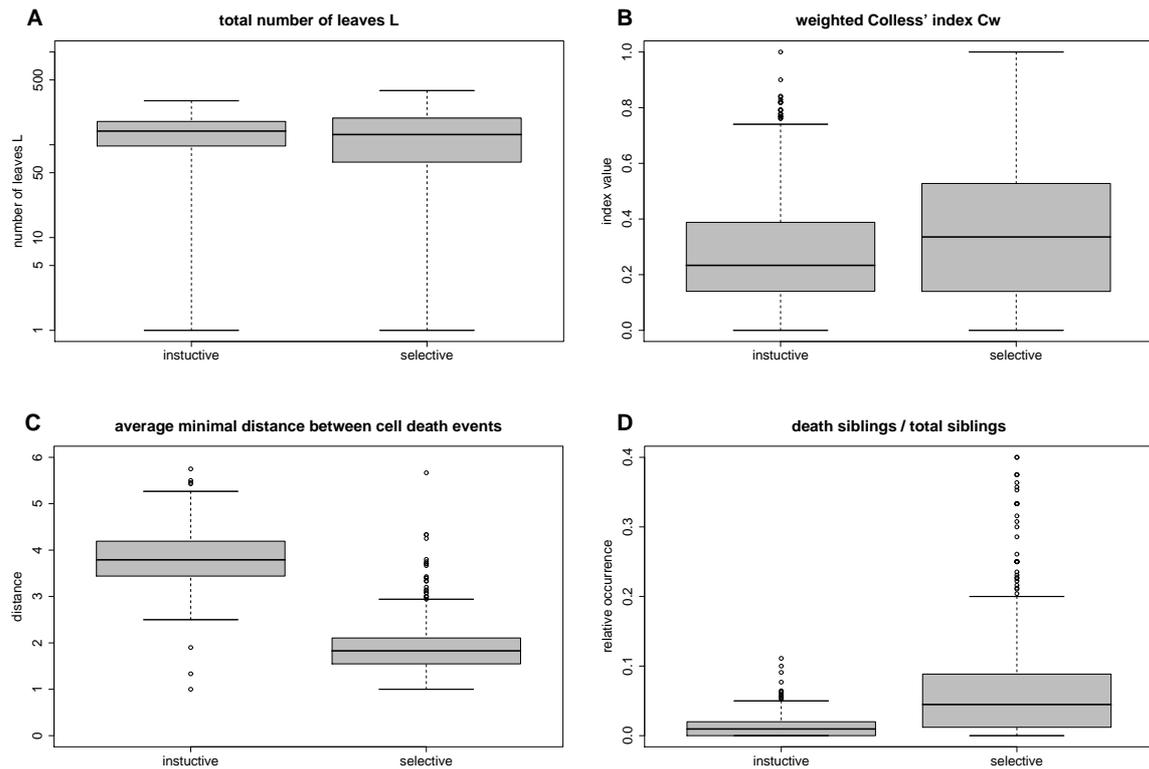


Figure 3. **Measures of tree shape.** Shown are boxplots of the distributions for the topological measures (A) total number of leaves L (shown on a logarithmic scale), (B) weighted Colless index C^w , (C) minimal distance between cell death events, (D) fraction of death siblings. Median values are shown by the thick bars, boxes correspond to the first and third quartile. Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

- [2] I. Roeder and R. Lorenz, "Asymmetry of stem cell fate and the potential impact of the niche observations, simulations, and interpretations.," *Stem Cell Rev*, vol. 2, no. 3, pp. 171–80, 2006.
- [3] S. J. Morrison, N. M. Shah, and D. J. Anderson, "Regulatory mechanisms in stem cell biology.," *Cell*, vol. 88, no. 3, pp. 287–298, 1997.
- [4] 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*, vol. 1106, pp. 30–40, 2007.
- [5] M. A. Rieger and T. Schroeder, "Exploring hematopoiesis at single cell resolution.," *Cells Tissues Organs*, Jan 2008.
- [6] I. Roeder and M. Loeffler, "A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity," *Exp. Hematol.*, vol. 30, no. 8, pp. 853–861, 2002.
- [7] 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*, vol. 12, no. 10, pp. 1181–1184, 2006.
- [8] I. Glauche, M. Cross, M. Loeffler, and I. Roeder, "Lineage specification of hematopoietic stem cells: Mathematical modeling and biological implications.," *Stem Cells*, vol. 25, no. 7, pp. 1791–1799, 2007.
- [9] I. Glauche, R. Lorenz, D. Hasenclever, and I. Roeder, "A novel view on stem cell development: Analyzing the shape of cellular genealogies.," accepted for publication in *Cell Proliferation*, 2008.