LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT
TECHNISCHE UNIVERSITÄT MÜNCHEN TUM

Helmholtz Zentrum München
Institut für Bioinformatik und Systembiologie

Diplomarbeit
in Bioinformatik

Models of murine embryonic stem cell dynamics on multiple scales

Bernhard Schauberger

Aufgabensteller: Prof. Fabian Theis
Betreuer: Dr. Carsten Marr
Abgabedatum: 28.04.2011

Ich versichere, dass ich diese Diplomarbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

28.04.2011 _____

Bernhard Schauberger

Abstract

The unique properties of embryonic stem cells (ESCs) and their potential usage in future medicine have raised the research endeavors worldwide. ESCs are capable of producing all cell types in the body of mammals and thus serve as a putative source for patient-specific cell therapies, alleviating cancer or eliminating the rejection of donated organs by breeding them *in vitro*. Despite several breakthroughs in deciphering the regulatory mechanisms of these cells, numerous secrets are far from being disclosed. The question of how they are capable of indefinite reproduction on the one hand and the creation of more than 200 functional cell types on the other hand has been treated by experimental and theoretical research. Laboratory experiments have created large amounts of data, and computational studies have further revealed valuable insights. Due to the complexity of the system, wholistic modeling is hardly applicable. Instead a multi-stage approach has been proposed, covering different aspects from noisy chemical reactions to system-level dynamic patterns. In this work, a Boolean regulatory network describing the first lineage decisions of a murine embryo is constructed of key transcription factors and its properties under different culture conditions analyzed. Additionally, the robustness against perturbations is assessed and the approach is compared to former studies. From this model, two biological hypotheses are conjectured and experimental strategies for their resolution are proposed. We propose a suppressive function of LIF on the inhibition of Nanog by Gata6 and Gcnf, and a similar function of Activin on the action of Erk, disallowing the suppression of Nanog and Oct4 by Erk if Activin is present. As the qualitative states of a Boolean network are not sufficient on a single gene scale, the heterogeneous expression of Nanog in murine ESCs is quantitatively analyzed with single cell time-lapse movies. An asymmetry in the allocation of Nanog molecules to nascent daughter cells and a substantial loss of Nanog during cell division are surveyed. The cell cycle times of the ESCs are shown to be independent from their Nanog expression level, and the existence of a slowly fading expression memory is revealed. With the help of a simple molecular model these features are studied as possible sources of the heterogeneity, revealing the high relevance of the asymmetric division. Additionally, former models of the heterogeneity in Nanog are validated against the data, indicating their possible inadequacy. The high impact of close experimental and theoretical collaboration is underlined with this work.

Zusammenfassung

Die einzigartigen Eigenschaften von embryonalen Stammzellen (ESCs) und ihre möglichen zukünftigen Anwendungen in der Medizin intensivierten die Forschung auf diesem Gebiet weltweit. ESCs besitzen die Fähigkeit, alle Gewebe und Zelltypen des Körpers in einem Säugetier zu bilden und sind deshalb eine geeignete Quelle für patientenspezifische Zelltherapien. Diese könnten beispielweise dazu dienen, Krebs zu behandeln oder durch spezifische Züchtung von Organen *in vitro* die Abstoßung von transplantierten Organen zu umgehen. Trotz zahlreicher Fortschritte im Verständnis der regulotarischen Zusammenhänge dieser Zellen sind noch viele Geheimnisse ungelüftet. Sowohl die experimentelle als auch die theoretische Forschung beschäftigt sich mit der Frage, wie Stammzellen einerseits unbegrenzte Selbsterneuerung und andererseits die Erzeugung von mehr als 200 Gewebetypen des Körpers intern ausbalancieren und regulieren. Experimente im Labor haben mittlerweile zu umfangreichen Datensammlungen geführt, und theoretische Studien haben in Zusammenarbeit damit weitere wertvolle Einsichten erbracht. Auf Grund der Komplexität des Systems ist eine vollständige Modellierung von Stammzellen im Computer nicht erreichbar. Stattdessen wurde ein Multiskalen-Ansatz vorgeschlagen. Die Idee dahinter ist die Abbildung von Vorgängen auf einer bestimmten Ebene des Systems, von verrauschten chemischen Reaktionen bis zu systemweiten dynamischen Vorgängen. In dieser Arbeit wird ein Boolesches Netzwerk mit zentralen Transkriptionsfaktoren erstellt, welches die ersten Differenzierungen in einem Mäuseembryo beschreibt. Dessen Eigenschaften werden under verschiedenen Kulturbedingungen analysiert. Zusätzlich wird die Stabilität des Netzwerks gegen Störungen gemessen und ein Vergleich mit früheren Studien gezogen. An Hand der Ergebnisse werden biologische Hypothesen aufgestellt und mögliche experimentelle Methoden zu deren Evaluation vorgeschlagen. Wir vermuten, dass LIF eine eindämmende Wirkung auf die Inhibition von Nanog durch Gata6 und Gcnf ausübt. Parallel dazu gibt es Hinweise auf eine ähnliche Funktion von Activin, das die inhibierende Wirkung von Erk auf Nanog und Oct4 aussetzt. Da die qualitativen Zustände in einem Booleschen Netzwerk auf der Ebene von einzelnen Genen nicht ausreichend sind, wird die nicht einheitliche Expression von Nanog in Mäuse-ESCs mit Hilfe von Einzelzellen in Zeitrafferfilmen quantitativ analysiert. Wir beobachten sowohl eine Asymmetrie in der Zuteilung von Nanog auf neu entstehende Tochterzellen als auch einen substanziellen Verlust von Nanog während der Zellteilung. Wir zeigen die Unabhängigkeit der Zellzyklenzeiten von der Menge an Nanog und die Existenz eines langsam verblassenden Expressionsgedächtnisses. Mit Hilfe eines einfachen molekularen Modells werden diese Eigenschaften als mögliche Quellen der Heterogenität analysiert und dabei der bedeutende Einfluss der asymmetrischen Zellteilung herausgestellt. Zusätzlich werden mit Hilfe unserer Daten bestehende Modelle der heterogenen Nanog-Expression evaluiert und deren mögliche Probleme dargestellt. Der positive Einfluss der engen Zusammenarbeit von experimenteller und theoretischer Seite wird durch diese Analysen hervorgehoben.

Acknowledgements

In thoughts succumbed and mentally so deep revolved
The sun has long yet left the midnight sky
All sentiments from dejected to exalted have within evolved
When a lonely cell I listen to but only hear a Celtic cry.

From populations to reactions and from networks to attractions
How to model what I see and elicit therefrom theory?
Thoughts alone yield no detections, but so does speaking with reflections
Carsten, you performed ideally, a thousand thanks to you from me.

The two leaders, filled with wit and heart and inspiration
The coders Michi, Jan and Michi for valuable contributions
Dominik and Daniel for mathematical and most precise intrication
Ivan and Christiane for curing all statistical illusions
I thank for their aid in this piece of science' creation.

The hackers Konstantin and Olli shall be honored for their computer skills
Adam, friend of Nanog shall be your name for rendering the cells so smart
Day by day and bead by bead Mashud with constant tracking all data fills
In short: ISF and CMB have been a pleasure right from the start.

With cordial thanks I do bestow all members of my family
Without their support there would not be this research diary.

To Katharina.

Contents

Chapter 1

Introduction

"Life is a study in contrasts between randomness and determinism" [139]. This citation defines the vital necessity of life to find a balance between both extremes, the inherently random nature of chemical processes on the one hand, and the execution of precisely defined actions on the individual or population level on the other. Since this task seems challenging and prone to errors, it is astonishing how tremendously complex life on earth can be. To unravel the very nature of this balance has been a desire of many researchers in the last decades, especially in the field of mammalian development. Developmental biology is engaged with the processes that define the way from a fertilized egg cell (zygote) to a fully grown individual, be it human, mouse or chimpanzee. On a cellular level, the pathways of development are already understood quite well (see [84, 158, 168] for reviews), but the subcellular mechanisms behind these ways are far from being conceived. A key role in this development is played by stem cells, subject of both laboratory and theoretical studies.

1.1 The importance of stem cells

A stem cell is defined as a cell with two pivotal properties: the ability to regenerate itself for an indefinite time and the ability to give rise to cells with different phenotypes.[1] The first action is termed self-renewal, while the latter is coined differentiation, referring to the separation of distinct functional lines of cells from one common progenitor. Stem cells are split into various subtypes, according to their origin and developmental potential. The first distinction is made between embryonic (taken from unborn embryos) and adult (taken from grown individuals) origin. The second distinction is defined by the number of different cell lineages a stem cell can generate. The least restricted cells are called totipotent (or omnipotent) stem cells, which are able to give rise to all organismic and extra-embryonic lineages. Usually only the zygote is totipotent. The second most potent stage form the pluripotent stem cells that can differentiate into all lineages of the body, but not extra-embryonic ones. A certain type of cells extracted from the Inner Cell Mass (ICM), a tissue forming during the early development of a mammalian embryo, is called embryonic stem cells (ESCs). These are pluripotent and give rise to all three germ layers (mesoderm, endoderm, ectoderm) that form the body and the germ cells, which are necessary for reproduction. More restricted are the multipotent cells that can produce offspring cells of usually more than one functional type of one lineage. An example are hematopoietic stem cells (HSCs) that produce all blood cell

[1] Definitions are leaned on [47, 76, 196].

types, or neuronal stem cells (NSCs) which can generate all cell types of the brain. The stem cells with the least potential are the unipotent ones, which give rise to exactly one cell type, like e.g. spermatogonial stem cells producing only sperm. Adult stem cells in general have less developmental potential than embryonic ones, since there are no adult omnipotent or pluripotent stem cells. Adult or embryonic progenitor cells (PCs) also have the ability to differentiate into other cell types of one or multiple lineages, but lack the long-term self-renewal ability of stem cells.
The first idea about the existence of stem cells was born in the 1950s after the atomic bomb drops over Hiroshima and Nagasaki, where many people died from degenerative diseases shortly after the disaster. The diseases primarily arose from the radioactive destruction of the hematopoietic stem and progenitor cells, impairing the production of fresh blood cells [184]. Some decades after this initial discovery, several other types of stem cells were identified and their putative medical potential has constantly increased. Nowadays they are considered as a great hope for curing heritable diseases, cancer or other genetic defects, for fundamentally eliminating the problems with blood and organ donations by breeding patient-specific artificial tissues, for *in vitro* drug screening, and also for understanding the essence of mammalian development [190, 193]. ESCs play a special role in this concert with their putative ability to generate whole tissues *in vitro*, which is already under clinical trial for several mesenchymal stem cells [59, 149], optic-cup formation [45] and others.[2] During the last decades, great progress has been achieved in the field of stem cell research, but many key questions still remain open.

1.2 Ethical considerations

Simultaneously with the successful derivation of human embryonic stem cells (hESCs) ethical issues concerning this research field were raised. Although ESCs (both murine and human) can be maintained and expanded *in vitro* for indefinite time, if appropriately cultured, the first isolation requires the killing of unborn organisms. This has not been a great issue for mice, but the more for humans, as many religious and ethical groups oppose to the usage of human embryos and hESCs in research or their genetic modification. They claim that life starts with fertilization of the egg, bestowing the right to life and physical integrity also on unborn potential beings. The result of the ongoing discussions is a separation of countries into countries that allow research with hESCs rather easily (e.g. Great Britain or Singapore), while others more or less severely restrict it (e.g. the USA or Germany). Nonetheless, the debates help to clarify the concept of human dignity and also shed light on the relation between possibly curing ill people on the one side and destroying embryos on the other side.
A recent breakthrough to bypass the ethical problems was achieved in 2006, when a Japanese group reported the derivation of induced pluripotent stem cells (iPSCs) from fibroblast (skin) cells [167]. The iPSCs do not require fresh embryonic tissues anymore, but cells taken from the skin of adult organisms and are then reprogrammed with defined factors (either proteins or chemicals) *in vitro*. This seminal study gave rise to the search for human equivalents, and in 2007 the same group reported the induction of human iPSCs from adult human fibroblasts [166]. Although the complete equivalence of iPSCs and true ESCs has been falsified in some

[2] See e.g. http://clinicaltrials.gov/ct/search?term=stem+cell&submit=Search for current clinical trials with stem cells.

aspects (e.g. chromatin structure [86] or gene expression [34]) and some of the reprogramming factors are actually oncogenes (that can trigger cancer), these issues are currently under focus to be resolved [68, 97, 189, 190] and a very active research community is dealing with improving the induction protocols and understanding the mechanisms behind it [65, 112].
The ethical controversies are continuing, and although a resolution might seem near at hand, a completely different approach could help to alleviate the concerns about genetic modifications of organisms and embryonic research: computer aided or theoretical stem cell research.

1.3 Previous studies about embryonic stem cells

In the following, the numerous experimental and theoretical studies about ESCs are shortly summarized.

1.3.1 Experimental studies

The ESC system is studied in many laboratories all over the world. The main model organism in the laboratory is the mouse (*Mus musculus*), a welcome research organism which is easy to breed and which develops rapidly with only three weeks from fertilization to birth. Mice and men share about 90% of the genome and many results from mouse are transferable to human [183], but there are often fundamental differences especially in the early development [193]. In particular, the ESC lines derived from both species differ in their signal response pathways and possibly their developmental stage, which resulted in 17 years of delay from the first derivation of murine ESCs (mESCs) [48, 107] to the first isolation of human ESCs [173].
Through numerable experimental studies, the transcription factor network behind pluripotency has been unraveled both in mouse [101, 132, 177, 181] and human [15] (see [31, 85] for combined studies), and the core triad of the genes Nanog, Sox2 and Oct4 has been established as the key regulatory entity in ESCs [127]. Additionally, epigenetic modifications like histone accessibility [10, 50, 55, 113] or methylation and acetylation [52] have been identified as essential for pluripotency. Only recently, Ronin was discovered as an essential epigenetic factor that acts independently of the established triad [43]. Regulation on the mRNA level of genes is performed by microRNAs, where ESCs have been found to contain a unique set of [56, 70, 169, 170, 188]. The roles of these microRNAs are diverse, from reducing noise over controlling the cell cycle to enabling rapid response to external stimuli in differentiation.
After a seminal publication [27] in 2007, which reported the heterogeneity of Nanog abundance in ESCs, several pluripotency-associated genes were found to be far from homogeneously expressed [66, 174]. These findings shifted the paradigm of a homogeneous ESC population towards a stable, but heterogeneous steady-state distribution of cells with different combinations of expression levels, defining pluripotency as a stimulus-sensitive ground state with promiscuous gene expression [61, 116, 123, 155, 160, 192, 196].
Several studies have delved into the role of noise in pluripotency and development, resulting in the acknowledgment of fluctuations as an essential driving force of development rather than a nuisance to be eliminated [46, 71, 99, 135, 141].
Altogether, the regulation of ESCs is understood at different levels, from the noisy expression patterns of single genes up to large deterministic transcription factor networks.

1.3.2 Computational modeling

The increasing knowledge about embryonic development and pluripotency of mammalian cells has also raised the need for theoretical models to cope with the growing amount of data. The ideal goal would be to model a cell with all its regulatory levels on the computer and to completely simulate experiments *in silico* instead of taking the burden of slow and error-prone laboratory work. However, even a single cell is tremendously complex and it will clearly never be possible to incorporate the different and intertwined regulations into one model. Thus computational simulations or theoretical analyses can only focus on certain aspects of regulation at a time. But these single aspects can often be combined to a large picture of regulation, which is then termed multi-scale modeling. Nonetheless, this has often lead to valuable insights and hypotheses or spared experimentators from potentially futile work.
A systemic perspective has been pursued and become standard in the modeling of pluripotency regulation, laid out in several studies which try to decipher the dynamics of transcription factor or protein networks in ESCs. An overview of the models described here and discussed in the following chapters is given in Table 1.1. For example, Chickarmane et al. [33] implemented the core triad of Oct4, Sox2 and Nanog and associated genes as a deterministic all-or-nothing switch (with ordinary differential equations, ODEs) that is able to capture the differentiation of embryonic cells into separate functional lineages (Primitive Endoderm [PE], Trophectoderm [TE] and pluripotent stem cells). A similar approach (with stochastic differential equations, SDEs) persecuted the cascade-like unidirectional differentiation of ESCs into three other cell lineages (osteogenic, chondrogenic and adipogenic) [104]. The authors also studied possible mechanisms to reprogram the terminally differentiated cells into pluripotent iPSCs, and found that amplification of the inherent transcriptional noise is sufficient for reprogramming.
The heterogeneity of protein expression levels inspired the construction of systems that reproduce this pattern by the incorporation of regulated fluctuations. In [78], the authors study an excitable system composed of only Nanog and Oct4. It tries to explain their experimental findings about the heterogeneity of Nanog and homogeneity of Oct4, but fails in vital aspects (see chapter 3 for details). The authors in [60] study two distinct systems that each are a possible explanation for the observed heterogeneity of Nanog. One model incorporates protein production noise as the only source of diversity, while the other utilizes an additional hypothetical protein and a negative feedback loop to produce oscillating expressions of Nanog. Two experimental strategies are designed to distinguish between the two models, one on population level and the other one on single-cell level.
An intriguing study showed that imperfect cell divisions with random allocation of cell contents to the daughter cells can result in exactly the same patterns of heterogeneity as transcriptional or translational noise, putting previous conclusions from experimental observations into question [73].
Theoretical descriptions of models for protein expression have been studied for almost a decade now, allowing, for example, the estimation of expression [90] or network [120] parameters or the distinction of diverse noise sources [165].
In [14], the author suggests a multi-scale approach to capture all features of cell regulation. On the one hand, highly detailed molecular models try to capture all essential chemical aspects of gene expression, from promoter status over transcription and translation to post-transcriptional or post-translational modifications, but cannot explain higher-level dynamics of individual cells. On the other hand, large-scale deterministic gene regulatory networks (e.g.

Authors	Model	Cell type(s)	Main aspect
Chickarmane et al. [33]	ODE	Murine embryo	Differentiation of zygote into four different lineages
MacArthur et al. [104]	SDE	mESC	Hierarchical differentiation of mESCs into three lineages
Kalmar et al. [78]	ODE	mESC	Fluctuating Nanog levels as an excitable system
Glauche et al. [60]	ODE	mESC	Two distinct models of fluctuating Nanog levels

Table 1.1: Theoretical models of mESC/embryonic properties and differentiation. For each model the mode of implementation is given in the second column, the organism and cell type studied in the third and the main aspect of this model in the fourth column.

Boolean networks) can account for cell behavior on a coarse level, but do not consider the details and inherent stochasticity or indetermancy of chemical reactions. Thus each model has its own scale and range of explanation. But in order to gain wholistic insights into cells, diverse levels have to be regarded.

1.4 Embryonic development

A scheme of the zygotic development of mouse is shown in Figures 1.1 and 1.2. In the first figure the development of a murine embryo until the late blastocyst stage is shown. In the second figure the schematic development of a murine zygote into all extraembryonic and embryonic tissues over numerous rounds of division and differentiation events is depicted. Each developmental stage or cell type has a unique set of genes that are expressed or not expressed (in the rounded boxes aside each stage), and many of the transitions between two different stages can be triggered by certain genes (denoted along these edges). The marker genes are not exhaustive, but represent a manually selected subset of genes curated from literature.
Note that the transfer of *in vitro* results from ESCs, which might be a cell culture-dependent artifact not occurring in normal development, to true *in vivo* events is an open issue. However, the study of differentiation and self-renewal mechanisms in ESCs can generate valuable insights and accumulate knowledge, which could be relevant for later application, e.g. in cell therapy.

1.5 Aim of this thesis

One of the questions currently addressed in stem cell research is how the pluripotent ESCs manage to tilt into the direction of either self-renewal or differentiation on demand. These processes are supposed to be under tight regulatory control [63] since they lay out the basis for all future developments of the organism and any disbalance could result in lethal damage. This thesis will deal with different levels of abstraction of this balance in mESCs. Two distinct systems are studied, both following current trends in research. A deterministic Boolean network incorporating defined culture conditions is analyzed in chapter 2. The expression of a single gene is studied and stochastic molecular models are analyzed in chapter 3. Implications

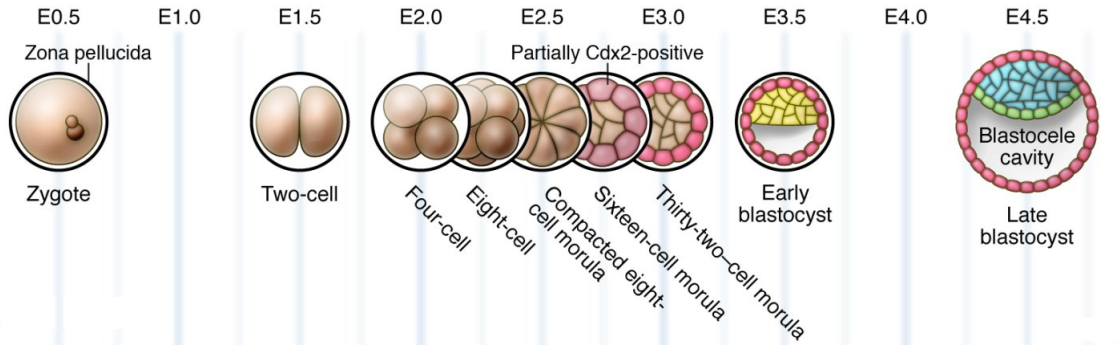


Figure 1.1: The development of a mouse embryo from the zygote (E0.5) until the formation of the late blastocyst (E4.5) after the implantation of the early blastocyst into the uterus. The numbers in the top row denote the days of embryonic (E) time after the fertilization. In the late blastocyst, the green cells represent the Primitive Endoderm (PE) tissue (Gata6 positive), the red cells the Trophectoderm (TE) tissue (Cdx2 positive) and the blue cells the Epiblast (Epi) cells (Oct4 and Nanog positive). In the early blastocyst, the yellow cells are the ICM cells (Oct4 positive). The image is taken and modified from [36].

for future research are investigated in both approaches.
The close collaboration with the Institute for Stem Cell Research at the Helmholtz-Zentrum München[3] is a vital aspect of this thesis since our results can directly be incorporated in the laboratory investigations there and some of the experimental propositions deduced from our models are already being tested in the laboratory. Thus both sides benefit from the close contact, which allows our models to be fit to real data (in this case, unique single cell time-lapse movies) but also helps the experimentators to gain further insight into the secrets of ESCs.

[3]http://www.helmholtz-muenchen.de/isf/haematopoese

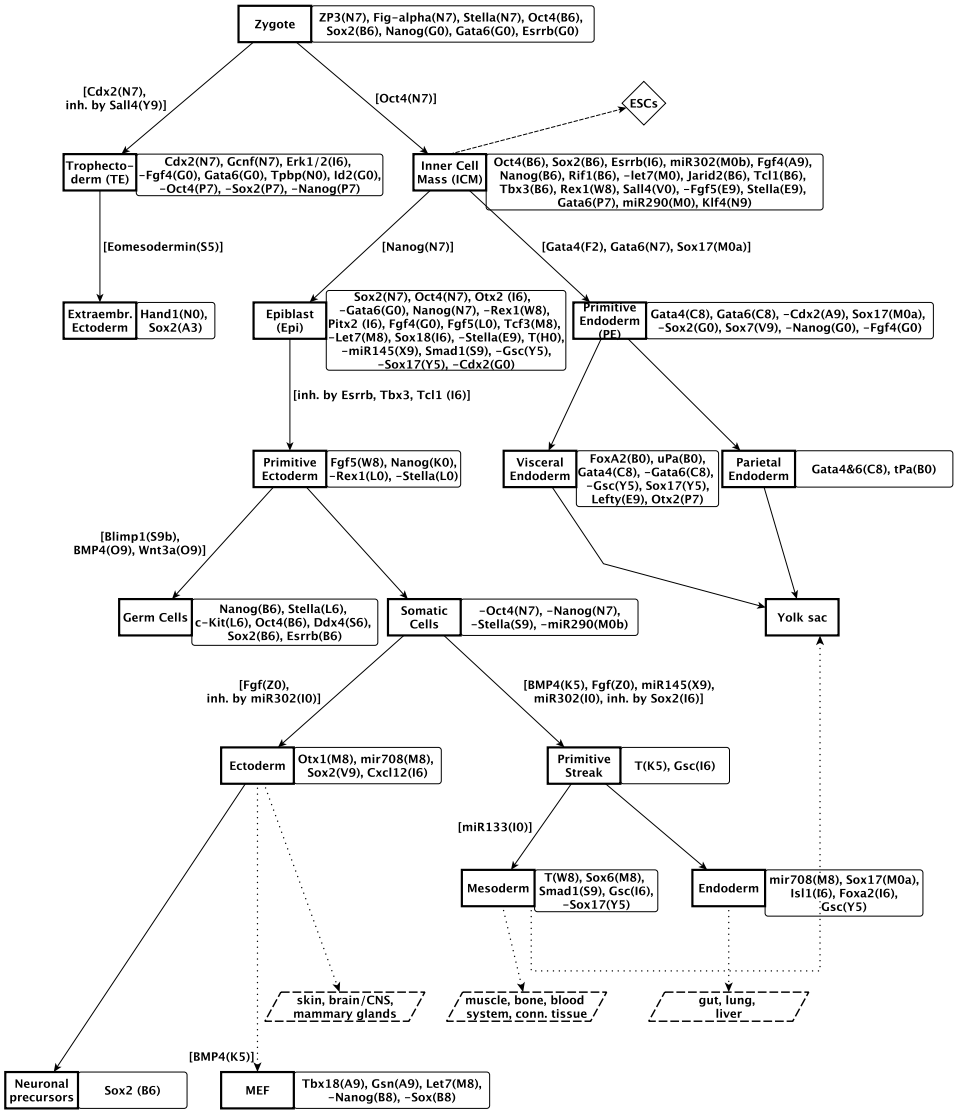


Figure 1.2: The developmental path of a murine zygote. Each developmental stage or cell type is associated with a distinct set of expressed or absent(-) genes, noted aside each stage in the rounded boxes. The codes (letter + number) after each gene reference to the study the expression state has been taken from. A map of these codes to the original studies is provided in Appendix 5.1. Transitions between stages can often be triggered by the expression or suppression(-) of certain genes, denoted by gene names along the edges. ESCs are derived from the Inner Cell Mass (ICM), but occur only transiently in the normal developmental path of mouse. The dotted lines summarize several differentiation steps into functional cell types.

Chapter 2

Boolean networks of embryonic development

2.1 Boolean networks

In this chapter, the applicability of Boolean networks to model the first stages of murine embryonic development is studied.

2.1.1 Background

To study dynamical properties of a gene regulatory network (GRN) on a large scale, a Boolean network can be used. A Boolean network is an ensemble of players (e.g. genes) which are either present (on) or not (off) with connections between them. These networks are named in honor of George Boole, an English mathematician, who invented the mathematics of logical reasoning. Despite their simple structure such networks can depict certain aspects of real processes very well. In a seminal publication from 1969, S. Kauffman proposed a random Boolean network and stated its applicability for biological questions [79]. Successful application in biology has been shown in several follow-up studies, for example analysis of T cell signaling [145, 185], segmentation of the Drosophila embryo [3] and the control of the mammalian and yeast cell cycle [49, 41]. A recent publication used a Boolean network to study the differentiation of hematopoietic stem cells from common myeloid progenitors into four mature blood cell lineages [92]. In [186] an extension of the two states of a Boolean network to multiple discrete states is performed and the biological relevance shown.
The main reason for studying a Boolean network is its capacity to model systems on a large scale and its lacking requirement for detailed kinetic parameters (which are often not known).

In this chapter a class of Boolean networks is set up to understand the first cell division and differentiation events in a developing mouse embryo: from the fertilized egg (zygote) to the formation of the 64-cell blastocyst with Trophectoderm (TE), Primitive Endoderm (PE) and Epiblast (Epi) cells. The latter two cell types are descendants of the Inner Cell Mass (ICM), as depicted in the upper part of Figure 1.2. The modeling process consists of two parts: in the first part we show that a Boolean network is sufficient to derive qualitative results that are consistent with the biological knowledge of the analyzed cell system. In the second part the Boolean network is extended to gain further insights into both the influences of various culture conditions and the development beyond the blastocyst stage. All networks are

validated against knowledge of murine embryogenesis. Robustness measures are calculated to test the stability of the networks, and testable biological hypotheses are derived. A workflow Petri net is shown in Figure 2.1.

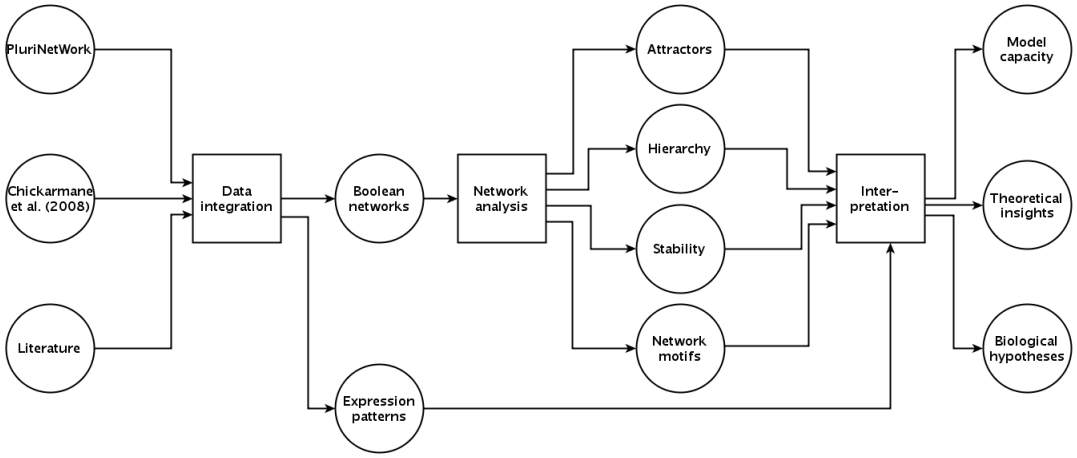


Figure 2.1: The Boolean analysis pipeline. The workflow is modeled as a Petri Net, a bipartite graph connecting states (data or results, in circles) with transitions (actions, in squares) [136].

2.1.2 Definitions and concepts

There are various ways to define a Boolean network [82]. In this work[1] a Boolean network $\mathbf{M}$ comprises a directed graph $\mathbf{G}$ with n nodes, in which each node i represents one gene[2] and has incoming edges from k_i other genes. A **state** $s(t) = (s_1(t), \ldots, s_n(t))$ of the network at time t is a vector of the states of all nodes at time t. The state of node i at time t, $s_i(t)$, is a logical (Boolean) function f_i of its input gene states at time $t-1$, $s_i(t) = f_i(s_{i_1}(t-1), \ldots s_{i_{k_i}}(t-1))$. The collectivity of all f_i is $\mathbf{F}$, and together with $\mathbf{G}$ it defines the Boolean network completely, $\mathbf{M} = (\mathbf{G}, \mathbf{F})$. The **state space** $\mathbf{S}$ of a Boolean network is defined as the set of all possible states of the network and consists of 2^n nodes. The **state transition graph** $\mathbf{T} = (\mathbf{T_G}, \mathbf{T_E})$ is a directed graph with the nodes $\mathbf{T_G} \subseteq \mathbf{S}$ and the edges $\mathbf{T_E}$ representing all allowed transitions in the network (a directed edge between two states $s(t)$ and $s(t+1)$ is present iff the transition from $s(t)$ to $s(t+1)$ is possible). It can be calculated (possibly for a given initial state) asynchronously, where exactly one of the possible updates in a certain state is chosen by some preference function, or synchronously, where all updates are performed at the same time.[3] In this work only asynchronous updates are used and the preference function is uniformly distributed, assigning equal probability to each edge. An **attractor** of the network is a state with no outgoing edges (excluding self-loops). This implies that once an attractor is reached it cannot be left anymore ($\forall i \in \{1 \ldots n\} : s_i(t+1) = s_i(t)$). The **basin** of an attractor is the set of all nodes that exclusively converge into this attractor. The **potential** of an arbitrary state is the set of attractors this state can fall into. Potentials and basins

[1] The notion of definitions is largely based on [91].

[2] In this chapter, genes, proteins and mRNAs are subsumed as single units (no distinction).

[3] Synchronous updates are of questionable use in biology, as many biological processes have an underlying stochastic nature, rarely enabling transitions to happen at the very same time.

can be used to analyze the *hierarchy* of the state transition graph. A shortest path between a given pair of nodes is defined as the path with the minimum number of traversed edges. From this, the **betweenness** of a node is the number of shortest paths that run through it [121]. Note that there are also other betweenness measures, e.g. based on random walks [12]. If a node in **T** has a high betweenness it possesses a possibly large destructive potential if knocked out [121], a putative indication of an important role in the network. In the following betweenness is measured in relative terms with values from 0 to 1, increasing from the lowest (0) to the highest betweenness (1) (within the given graph).

2.1.3 Capacities and limitations

A Boolean network is qualitative by nature. Qualitativeness implies clearly distinctive results, like existence statements, stability measures or hierarchies, with the input of only qualitative data. This is especially important with respect to biology because many data sources only provide qualitative information, namely expression or absence of a gene (but not its precise level) or the type of a regulatory interaction (but not the strength of the effect). The non-quantitative design allows for a parameter-free analysis, a property highly valuable when precise data is lacking. Additionally, a Boolean network is easily extensible with new nodes or interactions. Last but not least, the simple notion allows for the creation and analysis of large networks that would otherwise neither be analytically tractable nor interpretable, for example when modeled with differential equations or stochastic models [14]. This feature is extremely helpful in the light of systems biology, where functional modules and global dynamics are of interest rather than a fine-grained examination of individual genes.
Diverse questions can be addressed by a Boolean network, some of which will be dealt with in this thesis. First, one can analyze the attractors of the model and test whether these can be mapped to physiological cell states [72, 80, 81]. Second, the structure of the state transition graph can be interpreted (e.g. as the shape of Waddington's epigenetic landscape [179]). This might also unravel bottlenecks in the model, i.e. nodes that must be passed to reach a certain state and whose malfunction would destroy a substantial part of the system. Third, the stability of the network against perturbations can be analyzed, either with local perturbations (changing the state of a single node), or the global disruption of nodes or interactions. Finally, the influence of different choices of **F** on the model behavior can be analyzed.

Of course, the qualitative nature of Boolean networks is inadequate if quantitative results are desired. For example, in cells the precise protein level is often important (see e.g. [125, 131] for graded effects of Oct4 expression), which mostly cannot be captured by an on/off-switch. The molecular details of gene expression and other processes are neglected in a Boolean network, they are sacrificed for insights into global dynamics. Finally, there is no notion of time in a Boolean network, making it difficult to interpret intermediate states or trajectories in **T**.
Note that there are models with intermediate levels of complexity [14, 186] that try to incorporate both system-level features (meso-scale networks) and detailed dynamics of single genes.

2.2 Data sources

All input to our networks, from construction over analysis to validation, is extracted from previous peer-reviewed publications. This ensures that the networks incorporate the latest results, but also demands the integration of various sources of data into one model. The details of this integration are listed for each data source where it is relevant.

2.2.1 The GRN by Chickarmane et al.

In 2008, Chickarmane et al. [33] developed a model of the first two lineage decisions (between TE and ICM, and between Epi and PE) in the mouse embryo. Their model is based on a publication by Niwa [126] and is shown in Figure 2.3(a). This model is used as a seed for our Boolean network, as all genes and interactions are adopted.

2.2.2 The PluriNetWork by Som et al.

In 2010 G. Fuellen's lab published a comprehensive and highly informative network called the PluriNetWork [162]. This network is comprised of genes/proteins that play a role for either the maintenance or loss of pluripotency in mESCs. The network consists of 274 nodes and 574 interactions and is, to our knowledge, the largest systematic collection of protein interactions in mESCs when compared to other networks in mouse [187], or human [118]. It contains various types of proteins, including transcription factors, epigenetic modifiers and receptor proteins for extracellular signals, and three interaction types: activation, inhibition and interaction (unknown direction and effect). A gene only enters the network if it plays a role in pluripotency, evidenced in literature by an effect on mESC behavior. An interaction is only added if three criteria are fulfilled: (i) the interaction must be direct without intermediate agents; (ii) it must be analyzed for the mouse model; (iii) it must be involved in pluripotency. The curation is a manual process with weekly updates, currently based on 177 publications. The scope of the PluriNetWork, however, is limited if cellular events beyond pluripotency or other regulatory agents (e.g. microRNAs) are analyzed. The PluriNetWork is published as a Cytoscape [151] file. The authors demonstrate the applicability of the network with a tool called ExprEssence which identifies genes and interactions with the highest fold-change between two expression data sets [182].

2.2.3 Expression data

Gene expression levels are extracted from the literature, relying both on textual descriptions from Pfister et al. [137] (P), Niwa [126] (N), Silva et al. [156] (S) and Boyer et al. [16] (B), as well as on large-scale mRNA expression measurements from Aiba et al. [2] (A) and Guo et al. [62] (G) (the uppercase letters serve as a reference in Table 2.1). It needs to be stated that neither has the comparability of these studies been systematically analyzed, nor is it clear in general whether the mRNA levels correlate with the protein levels. Nonetheless, to validate our qualitative models, it seems a reasonable choice to make use of this simplified data set. Table 2.1 shows the expression states of all genes involved in the subsequent Boolean networks, used as a reference to compare network states with physiological cell states. If an explicit expression value (present or absent) for a gene/mRNA is found, a +/- is noted in the table. Two assumptions are made, namely the presence of Cdx2 in the zygote (due to a lacking upstream activator) and the expression of Erk in the Epi (as its activator Fgf4 is

present but no inhibitor). Additionally, all genes for which no expression value is found are assumed to be not expressed - which is reasonable in the light of lacking positive expression reports.
Expression/absence of a gene was considered evident if it is literally mentioned in one of the papers. For the transcriptional studies [2, 62] a twofold cutoff is applied to distinguish between expression and absence. In case of conflicts the literal description is assumed to be correct.

Gene	Zygote	TE	ICM/ESC	Epi	PE	Germ cell	Somatic cell
Oct4	+(B)	-(N)	+(B)	+(N)	-(A)	+(B)	-(N)
Sox2	+(B)	-(P)	+(B)	+(N)	-(G)	+(B)	
Nanog	+(G)	-(P)	+(B)	+(N)	-(G)	+(B)	-(N)
Cdx2	+(*)	+(P)	-(A)	-(G)	-(A)		
Gata6	+(G)	+(G)	+(P)	-(G)	+(A)		
Gcnf		+(N)	+(A)		+(A)		
Erk		+(A)		+(**)	-(A)		
Fgf4		-(G)	+(A)	+(G)	-(G)		
Rex1			+(A)	-(S)	-(A)		

Table 2.1: Gene expression table, showing expression (+) or lack (-) of mRNA in various cell types. If a value is missing, it is either not found in literature or unclear; in these cases, a '-' is assumed. (*) The putative expression of Cdx2 in the zygote is a model necessity, since no upstream activator of Cdx2 is included. (**) In the Epi state, expression of Erk is assumed as its activator Fgf4 is also present. Uppercase letters in brackets refer to the following papers: (P) = Pfister et al. [137], (N) = Niwa [126], (S) = Silva et al. [156], (B) = Boyer et al. [16], (A) = Aiba et al. [2], (G) = Guo et al. [62].

2.2.4 Embryonic development tree

A reduced version of the murine embryonic development tree shown in the introduction (Figure 1.2) is displayed in Figure 2.2. This part of the tree represents the first three lineage decisions of a fertilized mouse egg cell (zygote) and is used as a benchmark for the state transition graph hierarchy. The development starts with the zygote, and after four rounds of divisions (about three days), the 16-cell stage is reached where ICM and TE are already separated.[4] While the TE cells seam the outer layer of the embryo and are responsible for proper implantation of the embryo into the uterus and also its nourishment, the ICM is only an intermediate stage. Nonetheless, it is the source of pluripotent mESCs *in vitro*. With two further division events the 64-cell blastocyst stage develops where the ICM splits up into the Epi(blast) and PE cells. The Epi is still pluripotent as it can give rise to all cell types of the embryo (summed up with 'somatic cells') and germ cells. The PE is a cell layer that shields the Epi on the outer margins and provides both differentiation stimuli and nutrients.
It has to be noted that this tree contains some possible inconsistencies with literature knowledge. First, there is evidence that Gata6 is required, but not sufficient to drive PE formation [115]. However, this finding is contradictory to previous results [54], and for consistency with former models [33, 127] Gata6 is assumed to be sufficient for PE formation. Further, Nanog is

[4]The developmental stages can be found in detail in [62, 84].

assumed to enable the distinction between germ and somatic cells, as it is expressed in germ cells [16] but not in somatic cells [126]. Yet there is no evidence that Nanog levels actually drive these different fates.

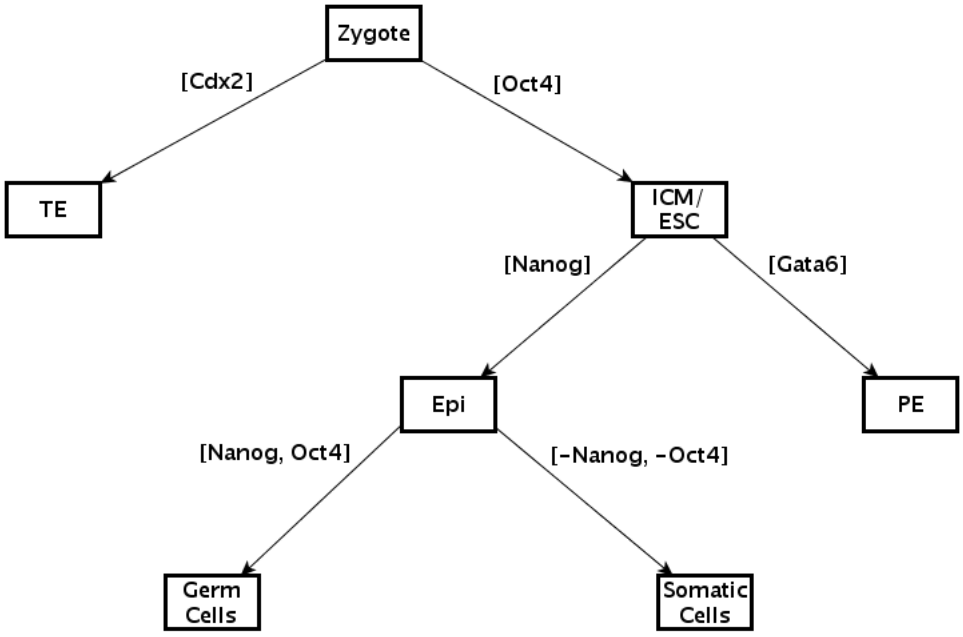


Figure 2.2: The upper part of the murine development tree, used as a reference for assessing the hierarchy of the state transition graphs. The genes written on the edges are both required and sufficient to form the corresponding cell type below (with the exception of somatic cells where the pluripotency genes Oct4 and Nanog have to be switched off). For references to these differentiation events, please refer to Figure 1.2. The expression signature for each cell type can be found in Table 2.1. Primitive Ectoderm and Epiblast are combined to one node (Epi) as their expression profile is equal with regard to the genes in Table 2.1.

2.3 A Boolean network of murine blastocyst creation

In this section a Boolean network is constructed from the data sources in 2.2 and analyzed with regard to the concepts above.

2.3.1 Network construction

For the network construction the GRN published in [33] (shown in Figure 2.3(a)) is used as a network base by adopting all genes and interactions. The interactions missing in that model, but present in the PluriNetWork, (see below for details) are included in our network. Our network topology can be seen in Figure 2.3(b). For a detailed explanation of each protein's role in embryonic development and the reason why it is included in the network, please refer to Appendix 5.2.
The logical rules (shown in Table 2.2) are the same for all proteins, with one inhibitor sufficient to switch off a target gene (in short 'one inhibitor suffices'). Activators act in a non-exclusive

and separate fashion, and so do inhibitors. Thus a protein is present if at least one activator and no inhibitor is present. The only exception from this rule is the heterodimer of Oct4 and Sox2, which is known to form when both proteins are present [32]. In the following this Boolean network is referred to with the shorthand CME (Chickarmane Model Enhanced).

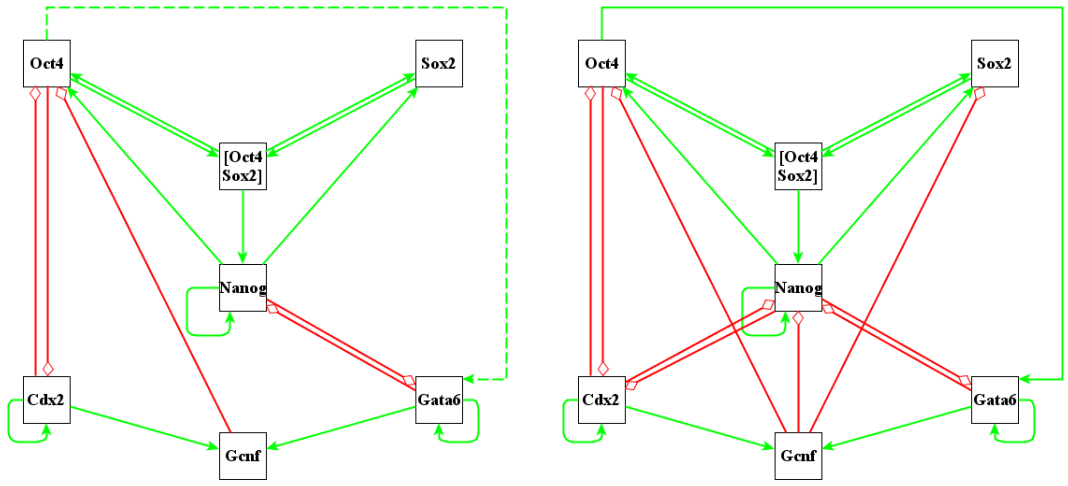


Figure 2.3: Two GRN topologies based on the ODE model by Chickarmane et al. [33] and the PluriNetWork [162]. Each box represents one gene, green arrows denote activation, and red (diamond) arrows denote inhibition. **(a)** The network with the interaction structure as in [33]. The dashed activation from Oct4 to Gata6 is a hypothesis by the authors in [33]. **(b)** The CME network, an extended version with the originally missing interactions included. These are the mutual inhibitions between Cdx2 and Nanog, and inhibition of Gcnf on both Nanog and Sox2.

$$\begin{aligned}
\text{Oct4} &= (\text{O4S2} \vee \text{Nanog}) \,\&\neg\, (\text{Cdx2} \vee \text{Gcnf}) & \text{Sox2} &= (\text{O4S2} \vee \text{Nanog}) \,\&\neg\, (\text{Gcnf}) \\
\text{Nanog} &= (\text{O4S2} \vee \text{Nanog}) \,\&\neg\, (\text{Gata6} \vee \text{Gcnf} \vee \text{Cdx2}) & \text{O4S2} &= (\text{Oct4} \;\&\; \text{Sox2}) \\
\text{Cdx2} &= (\text{Cdx2}) \,\&\neg\, (\text{Oct4} \vee \text{Nanog}) & \text{Gata6} &= (\text{Gata6} \vee \text{Oct4}) \,\&\neg\, (\text{Nanog}) \\
\text{Gcnf} &= (\text{Cdx2} \vee \text{Gata6})
\end{aligned}$$

Table 2.2: Logical functions for each gene in the CME (Figure 2.3(b)). O4S2 serves as a shorthand for the heterodimer Oct4-Sox2. $\vee$ is a logical OR, & is a logical AND, and $\neg$ is a logical NOT.

2.3.2 Analysis of the state transition graph

The state transition graph $\mathbf{T}_{CME}$ calculated with a zygote as starting state (see Figure 2.4 for a compressed version) is acyclic, implying a clear developmental direction. The graph maps well to the development tree in Figure 2.2. How can we measure the quality of this bijection? First, the attractors (yellow rectangles) need to correspond to true cell types. This is the case for all attractors (listed in Table 2.3). These attractors can be mapped onto physiological cell

states, a technique which has been established in numerous publications [79, 80, 81, 72]. The attractors could also be interpreted as the terminal basins at the bottom of Waddington's epigenetic landscape [179]. The PE attractor corresponds exactly with the expression values in Table 2.1 for the physiological PE state. The Epi state matches with both pluripotent states above (ICM/ESC and Epi), but a lacking distinctor (e.g. Rex1) leaves us with uncertainty about which state it exactly is. In the following, we will assume this state to match with the Epi state. The all-off state can either be considered a modeling artefact (a gene can go off if no activator is present) or the true somatic state (for which no positive marker is included). The (TE) state does not completely match with the TE expression values in Table 2.1 since Gata6 is missing. This is not the case in the TE state where Gata6 is also expressed, now clearly marked as TE. The lack of Gata6 in the (TE) state is due to a lacking activator of Gata6 if Oct4 is off, which thus allows for the formation of a TE-like attractor without Gata6. Second, to recover the hierarchical character of the embryonic development tree (Figure 2.2) there should be an intermediate state that maps to the ICM, if we assume the second attractor to be Epi (see below). This ICM state is characterized by its potential to give rise to the Epi, PE and somatic state but not the TE state. Such a state is indeed present, marked in light green. This state has also the highest betweenness (1.0) of all states (range of the other states is 0-0.8), suggesting a highly important role in development because most paths to the PE, Epi and somatic state lead through this state.
It has to be noted that some facts do not fit well to the tree, namely one node (the lower one with three members) can fall into either Epi, (TE) or somatic cell, but not PE. So it avoids any known physiological interpretation since it does not match with any possible intermediate state of the development tree. However, the connectivity of this node is low and only one edge leads to the Epi state, suggesting a modeling artefact. Additionally, different on/off-switching dynamics of the genes might play a role in the decision for one or the other lineage (namely Nanog turning off faster than Cdx2 in this case) in a real cell.

Gene	Starting state	Attractors				
Oct4	1	0	1	0	0	0
Sox2	1	0	1	0	0	0
Oct4-Sox2	1	0	1	0	0	0
Nanog	1	0	1	0	0	0
Cdx2	1	0	0	0	1	1
Gata6	1	1	0	0	0	1
Gcnf	0	1	0	0	1	1
	Zygote	PE	Epi	somatic	(TE)	TE

Table 2.3: The zygotic starting state (second column) and the attractors of the CME (Figure 2.3(b)) in columns 3-7. The dimer Oct4-Sox2 is assumed to be present in the zygote, as both Oct4 and Sox2 are present, too. A possible physiological interpretation of each attractor is given in the last row; for details please refer to the text.

2.3.3 Stability

An interesting feature of the attractors and the network itself is its robustness against perturbations. To quantify both the global and local stability, two methods are designed and

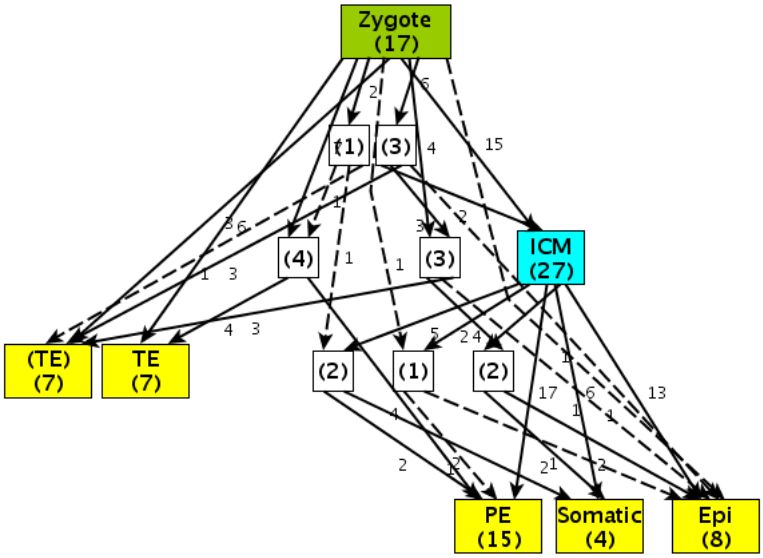


Figure 2.4: Compressed state transition graph $\mathbf{T}_{CME}$ for the CME (in Figure 2.3(b)) with a zygote as starting state. Compression is performed by grouping nodes with the same potential into one supernode, and edges connecting the same supernodes into one superedge. Numbers in the nodes indicate the true node count that is compressed into one. Edge numbers represent the edge count compressed into one, respectively. The uncompressed graph contains 101 nodes and 278 edges.
The yellow states denote the attractors with physiological interpretation, while the green one at the top is the zygotic starting state. The light blue state (27) represents the putative ICM state. The dashed edges all have weight 1 (singular edges also in the original state transition graph).

applied. The global stability is measured by removing single nodes or edges and checking $\mathbf{T}_{CME}$ for changes. A systematic knock-out of each gene and interaction is performed; the results are listed in Tables 2.4 (genes) and 2.5 (interactions). For each disruption the number of missing (M) and new (N) attractors in comparison with the original attractors is counted. If the removal of one component changes the set of attractors, it is considered essential. However, there are different levels of essentiality, indicated by the gravity of the disruption effects. The knock-out of Nanog disrupts three attractors, underlining the high importance of Nanog for murine embryonic development. The interactions whose failure causes the most problems are the auto-activations of Cdx2 and Gata6. The removal of these self-stimulating loops disables the formation of either TE or PE, alluding to the relevance of these genes in those tissues. Interestingly, the removal of the auto-activation of these genes yields the same effects as the knock-out. This underlines the necessity of auto-activation for creating regions of bistability [9, 175]. The destruction of the self-binding site in the promoter of these genes is a possible method to test this experimentally: if the tissues do not form, the auto-activation is confirmed to be required.
The local stability, or stability of the attractors, is measured differently. In every attractor, b bits of its state vector are flipped, and subsequently the newly reachable attractors from this perturbed state are calculated. A bit is the expression state of a gene. The bit flips could be interpreted as expression noise or transient down-turn ($1 \rightarrow 0$) or overexpression ($0 \rightarrow 1$) of a gene. To measure stability quantitatively one can count the number of bit flips that end up

in exactly the same attractor, then those that end up in either the same attractor and some more, and those that end up in completely different attractors (a related measure of stability has been introduced in [89]). We devised a simple scoring scheme with assigning two points to an unchanged attractor ($p_i = 2$), one point to a mixture ($p_i = 1$) and zero points to completely different attractors ($p_i = 0$). The number of possible simultaneous bit flips is $m = \binom{n}{b}$ and the total sum $P_b = \sum_i^m p_i$ is used to calculate $I_b \in [0, 1]$ with $I_b = \frac{P_b}{2*m}$. The higher I_b, the more stable an attractor is considered, as it is less prone to changes if one protein alters its expression. This measure is calculated for flips of 1 and 2 bits, and the results are given in Table 2.6. It can be seen that I_1 and I_2 rank the attractors in an equal fashion, indicating the consistency of the measure. The numbers tell us that Epi is the most stable attractor, although the stability of any attractor is not far apart from the rest. The high stability of the pluripotent state may be due to its gene expression similarity with the zygote, yielding a higher chance to fall into any attractor after a perturbation, thus also itself. Additionally, this is in line with the idea of a pluripotent 'ground state' proposed in [155, 123]: pluripotency is not a state that needs constitutive enforcement from outside, but rather a passive stable state that needs extrinsic signals to be exited. The PE attractor state is the second most stable, possibly generated by its shared ancestry with the Epi state. The equal stability of the TE and (TE) attractors underline both its similarity and our interpretation of equating (TE) with TE. The least stable attractor is the somatic state, which can be explained by its lack of any positive marker gene, rendering it susceptible to small changes.

KO Gene	M	N	Comment
Oct4	-	-	(no changes); loss possibly compensated by Nanog and Sox2
Sox2	-	-	(no changes); loss possibly compensated by Nanog and Oct4
Nanog	3	-	Epi, (TE) and somatic state are missing; consistent with the high importance of Nanog in pluripotency and development [26, 27, 114, 155, 16, 126]; the missing (TE) state could be explained by a required intermediate state with Nanog
Cdx2	2	-	(TE) and TE are missing; consistent with the requirement of Cdx2 for TE [128]
Gata6	2	-	PE and TE missing; consistent with Gata6 expression in both tissues [22, 62]
Gcnf	1	1	The somatic state is missing, possibly a modeling artefact; a mixture of Epi with PE is an attractor now, consistent with the pluripotency-silencing role of Gcnf in PE [126, 117]

Table 2.4: Global stability as knock-out analysis of genes in the CME network. For each change the state transition graph with a zygotic starting state is calculated, and the attractors are compared with the original ones. The first column shows the knocked-out gene, column 2 lists the number of missing attractors (M), and column 3 the number of new attractors (N). In the comment column the changes are listed in detail for each knock-out.

2.3.4 Role of the Gata6 activation by Oct4

In [33], the authors suggest an activation of Gata6 by Oct4. This hypothesis is based on the knowledge that on the one hand Oct4 activates the pluripotency machinery with Oct4, Sox2

KO Interaction	M	N	Comment
Nanog → Oct4	-	1	Epi without Oct4 is new; this is not pluripotent by definition [23]
Nanog → Sox2	-	1	Epi without Sox2 turns up; this is not pluripotent by definition [51]
Cdx2 → Cdx2	2	-	Both the TE and (TE) states are missing; consistent with the high relevance of Cdx2 in the formation of TE [128]
Nanog $\dashv$ Gata6	3	-	The three missing states are Epi, somatic and (TE); this indicates the relevance of Gata6 suppression for the formation of these tissues
Gata6 → Gata6	2	-	Missing are PE and TE; consistent with the expression maintenance of Gata6 in both tissues [22, 62]
Cdx2 → Gcnf	1	1	The (TE) state has no Gcnf expressed
Gata6 → Gcnf	1	2	The PE state has no Gcnf expressed, and an Epi-like attractor, but without Nanog and with Gata6 is new - this is possible as Gata6 does not turn off the pluripotency genes Oct4 and Sox2 by itself.

Table 2.5: Global stability as knock-out analysis of interactions in the CME network. For each change the state transition graph with a zygotic starting state is calculated, and the attractors are compared with the original CME attractors. The first column shows the knocked-out interaction (A → B denotes activation of B by A, whereas A $\dashv$ B denotes an inhibitory function of A on B), column 2 lists the number of missing attractors (M), and column 3 the number of new attractors (N). In the comment column the changes are listed in detail for each knock-out. Only those disruptions that actually cause the attractor set to change are listed.

Attractor	I_1	I_2
Epi	0.71	0.55
PE	0.64	0.48
TE	0.57	0.33
(TE)	0.57	0.33
Somatic	0.43	0.26

Table 2.6: Local attractor stability of the CME (Figure 2.3(b)). The higher I_b, the more likely the attractor will return to itself after a perturbation of $b = 1$ or $b = 2$ bits, respectively. The Epi state is the most stable, while the somatic state is the least stable in both cases.

and Nanog [101], while on the other hand an excess of Oct4 triggers PE formation [125, 131]. This suggests that intermediate levels of Oct4 activate Nanog while high levels repress it. One possible mechanism generating this pattern is the activation of Gata6 (which is known to repress Nanog) by Oct4 (another mechanism is analyzed in [78]). However, in our CME network, the knock-out of this activation yields no visible effect if the starting state of **T** is the zygote from Table 2.1. But when the starting state used in [33] (without Gata6) is applied to the CME network, this activation is indeed required for the formation of PE. As there is no activator of Gata6 it would never be turned on if not present right from the start.

So we conclude that the hypothetical activation of Gata6 by Oct4 is not a necessity for PE formation in our Boolean network, but nonetheless it is kept in the CME to enable a better comparison of our network with the ODE approach in [33] (see below).

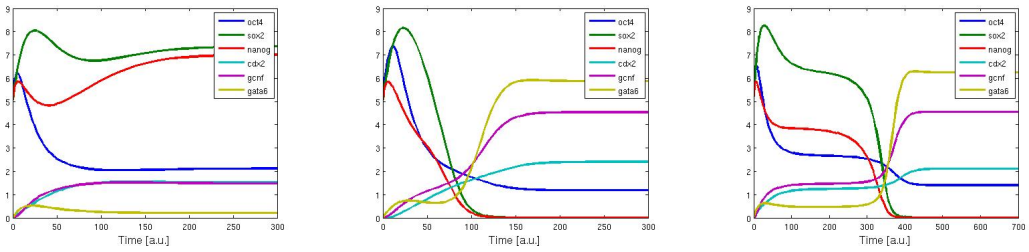
2.3.5 The ODE approach by Chickarmane et al.

The network shown in Figure 2.3(a) is originally implemented in [33] with ordinary differential equations (ODEs). These describe the change of concentration over time for each protein, depending on the current state of its promoter (no difference between genes and proteins in their model) and the concentration of its associated regulators. Chickarmane et al. model the time courses of six transcription factors (TFs), starting from a defined state and for different levels of an external stimulator of Oct4. With this setting they obtain three steady states of the model, which they assign to TE, PE and 'stem cell' (SC). Additionally, they study the switching behavior of an antagonistic gene pair, Nanog and Gata6. From their model they derive two hypotheses, namely Oct4 activating Gata6 and a heterodimer of Oct4 and Gata6 repressing Nanog instead of Gata6 only.
This model is able to explain biological facts (see section 2.3.6 for details), but their approach contains two serious problems. First, the six stated players require 44 modeling parameters (basal transcription rates, decay rates and interaction strengths), all of which are chosen without proper biological motivation. This makes the model very susceptible to subtle modifications, as shown in Figure 2.5(c), where the system falls into PE rather than SC when the parameter b_4, the binding strength of the heterodimer Oct4-Cdx2, is halved. Second, four interactions are missing, two of which have been reported before the time of publication (see Figure 2.3 for a comparison). After incorporating these interactions, the ODEs showed a completely different behavior (Figure 2.5(b)), resulting in PE instead of SC under the same starting conditions. This change already occurs at very low rates of inhibition (0.5/time unit), compared to the values of the already included inhibitions (between 5/time unit and 15/time unit). Taken together, this would imply a very unstable regulation of embryonic development, which is not likely, as the critical process of embryogenesis is supposed to be under tight regulative control and to some extent buffered from environmental noise [63].
A similar approach as in [33] has been analyzed in [104], which essentially suffers from the same problems.
Summing up, the results of the parameter-based ODE approach are very hypothetical for large systems without measured rate parameters as all conclusions are based on thin ground.

2.3.6 Result comparison between the CME and the ODE approach

Although the high level of detail of an ODE model is probably of importance to model complex biology, the relatively coarse Boolean approach is capable of reproducing the relevant ODE results, as well as additional results. What matters, however, is not the comparison of both modeling approaches, but how well embryonic development can be explained by each. In the following some essential biological aspects are considered, and for each it is stated whether the Boolean network CME and/or the ODE model from [33] captures this aspect.

Levels of Oct4 expression are known to be decisive for the distinction between PE, TE and ESC [125, 131]. This fact is not captured by our Boolean network. However, modeling precise gene levels is not impossible, as shown in Appendix 5.3. What we can state is that no TE can be formed if Oct4 is present (data not shown). For the same reasons we can neither



(a) (b) (c)

Figure 2.5: Time courses for the ODE model in [33] (x-axis: time in arbitrary units, y-axis: gene/protein concentration). Each line denotes the concentration of one gene over time. The flat trajectories at the end represent a steady state of the model, which can possibly be mapped on a physiological cell state. **(a)** Time courses obtained from the original equations and parameter setting as in [33] **(b)** Time courses after inclusion of the omitted interactions (taken from the PluriNetWork). Inhibition strengths are moderate in comparison to the original values **(c)** Time courses with original interactions only, but one parameter changed ($b_4 = 5$). Note that there is no steady state yet at time 300, which occurs only later. For all time courses the value of the external signal A equals 10.

confirm nor reject the hypothetical activation of Gata6 by Oct4 and the putative heterodimer of Oct4 and Gata6 suppressing Nanog (rather than Gata6 alone). Up to now there has been no evidence for both hypotheses (activation and dimerization) in the literature.

The mutual inhibition of Nanog and Gata6, as well as Cdx2 and Oct4, along with the auto-activation of these genes, creates bistability in both our Boolean and the ODE model. This bistability is of importance in embryonic development to foster distinction between lineages [47], and the two mentioned motifs reflect the first two lineage decisions in the development tree: Oct4 *vs.* Cdx2 decides about ICM or TE, and Nanog *vs.* Gata6 decides about Epi or PE. If the two genes of such a motif are expressed simultaneously, the decision between them is a random process and the cell fate can be defined by either winner. This is evidenced by an inspection of $\mathbf{T_{CME}}$, in which all states with two antagonistic genes expressed at the same time are not stable and can fall into either of two trajectory sets.

The explicit modeling of time courses with ODEs is a great advantage of this model type over Boolean networks, which have no notion of time. However, as stated above, interpreting the temporal development of expression is only reasonable if the input parameters are taken from biological experiments. Thus the authors in [33] only consider the steady states of their model. In contrast, in our CME network the time course of expression is visible in the hierarchy of the state transition graph, which contains a clear direction from the zygote to differentiated states. This hierarchy is a strength of the Boolean network, where it is implicitly incorporated in the network structure and the logical functions of the nodes. Here, the terminally differentiated nature of the attractors is shown by their attractor property and the acyclicity of the state transition graph. Additionally, developmentally important intermediate states can be found along the trajectories from the zygote to the final cell states. This novel finding is (to our knowledge) not yet published and resembles the results in [92], suggesting a general incorporation of differentiation hierarchy in the transcription factor network.

The robustness of the GRN has been assessed with our network on both local and global scale. The results underline the known physiological function of either genes or interactions. In the ODE model this has not been treated. In the publication by Chickarmane et al. [33], four inhibitions that are present in the PluriNetWork are missing. These are (i) the mutual antagonism between Cdx2 and Nanog [29] and (ii) the inhibition of both Nanog and Sox2 by Gcnf [126]. The network topology without these inhibitions can be seen in Figure 2.3(a). To study the roles of these interactions a simultaneous knock-out of all four of them is performed and the state transition graph calculated. The attractor set of the disrupted model is extended by one attractor which is a mixture of the TE and Epi states. This state is not consistent with any cell state from Table 2.1 since it expresses both pluripotency (Sox2, Nanog) and TE (Cdx2, Gcnf) markers. Although this could be interpreted as an intermediate state between ESC/Epi and TE, it can be seen in the state transition graph that this state originates if a cell begins to turn into TE (switching on Cdx2 and switching off Oct4), but fails to shut down Nanog and Sox2, since no inhibitor of them is present. Thus one can consider this state to be unphysiological. Additionally, a mapping of the structure of $\mathbf{T}$ onto the development tree fails because there are some states with the potential to fall into TE or PE, but not Epi, and also states that converge into either TE or Epi, but not PE. These states cannot be mapped to any differentiated or intermediate state within the tree.
Taken together, we can conclude that the missing interactions are of biological importance in the network. The two inhibitions from Gcnf to both Nanog and Sox2 are obviously necessary to finally switch off the pluripotency genes in the TE, while the mutual inhibition between Cdx2 and Nanog could be of importance for a fostered distinction between ICM/Epi (where Nanog is expressed and Cdx2 is lacking) and TE (where Cdx2 is expressed and Nanog is lacking).

In [33], the authors claim that the reprogramming from PE to the SC state is better achieved with overexpression of Nanog than by suppression of Gata6. We confirm this finding since the suppression of Gata6 in the PE attractor yields the all-off (somatic) state, whereas constitutive Nanog expression enables reprogramming to Epi.
In the ODE model there is a steady state which the authors name 'Differentiated Stem Cell'. This state shows markers of TE (Cdx2, Gata6, Gcnf), but Oct4 is also expressed, and they do not explain what physiological state this should represent. In our network this possibly unphysiological state does not occur.
Additionally, we confirmed the known master roles of Nanog and Oct4 in pluripotency. In our network, ESC/Epi cells with only Nanog off are prone to (but can revert to Nanog on), while ESC/Epi cells with only Oct4 off are determined to differentiate into either PE or somatic cells (data not shown). This is consistent with the role of Nanog as a 'safeguard' [27], and Oct4 as a marker of pluripotency [23].

2.3.7 Summary

In this section we have shown that a Boolean network is able to capture the essentials in murine embryonic development. Additionally, it has been shown that for qualitative inputs a Boolean network is at least as informative as an ODE model, while being robust against parameter changes (as there are none). In the next section the CME network constructed here is extended to give a deeper insight into the development of murine embryos under certain culture conditions.

2.4 An extended network with culture

In the previous section we have shown that Boolean modeling is sufficient for qualitative insights into murine development from the zygote to the 64-cell blastocyst. Now our aim is to gain further insights into the behavior of murine embryonic stem cells *in vitro* under certain established culture conditions used every day in the laboratory to maintain and proliferate different types of stem cells.

2.4.1 Culture conditions and network construction

First we delve into the role of LIF (leukemia inhibitory factor), which together with a specific (fetal calf) serum is used to maintain mESCs *in vitro* for unlimited time and passages [30, 42, 161]. The serum contains nutrients and, among others, the protein BMP4 (bone morphogenetic protein 4) which is necessary to maintain pluripotency. The second medium of interest is Activin, used together with FGF2 (fibroblast growth factor 2) and another serum to arrest mEpiSCs (murine Epiblast stem cells) in their current state [18, 42, 176]. In the following, mEpiSCs and Epi(blast) are used interchangeably.

The construction of the Boolean network with culture conditions, in short NCC, starts with the adoption of the CME network, shown in Figure 2.3(b). The hypothetical activation of Gata6 by Oct4 is not included since there is no literature affirmation for this interaction. The two culture media of interest, LIF + serum (in short LIF only) and Activin + FGF2 + serum (in short Activin only), are added and the six interactions of these proteins with the other players are taken from the PluriNetWork [162]. In order to distinguish between mESCs and mEpiSCs the protein Rex1 (also known as Zfp42, zinc finger protein 42) is added since it is known to be a marker of mESCs that is tuned down when differentiating into Epiblast [18]. This downregulation is known to be mediated by FGF4, activating Erk1/2 (extracellular signal-regulated kinase) signaling [18, 96]. Therefore we include these two proteins as well. As previously the activations and inhibitions between these three genes are taken from the PluriNetWork. Four further interactions are added that are not present in the PluriNetWork. These are the inhibition of Rex1 and Oct4 by Erk, evidenced in [94], the inhibition of Nanog by Erk, taken from [28], and the inhibition of Erk signaling by LIF, as shown in [96]. The final network is shown in Figure 2.6.

In section 2.3 it has been shown that rather simple logical rules, where one inhibitor is sufficient to stop the expression of a gene, are adequate to model embryonic development. Here we follow that approach, with the exception of the Oct4 and the Nanog rule. It is known that LIF and Activin activate Nanog and Oct4, while Gata6 inhibits Nanog and Gcnf and Erk inhibit both Nanog and Oct4. If the 'one inhibitor suffices' rule is adopted for Nanog, ESCs cannot be stably maintained because even in the presence of LIF, PE can still be produced from ICM cells. An inspection of the state transition graph for that network reveals that this is due to the stochastic switch between Nanog and Gata6 in the ICM state. If Gata6 and Gcnf are off, the PE state cannot be reached. Thus we assume LIF to exert an activation effect on Nanog that is stronger than the suppression effect of Gata6 or Gcnf. Furthermore, the suppression of Nanog and Oct4 by Erk signaling impairs the maintenance of Epi cells in the presence of Activin if the default rule is applied. Thus we also assume Activin to activate Nanog and Oct4 even in the presence of Erk. These rule modifications are hypotheses only, as no literature evidence is found for such activation effects. Possible strategies to test this hypothesis experimentally include a promoter inspection of Nanog and Oct4 (do they contain

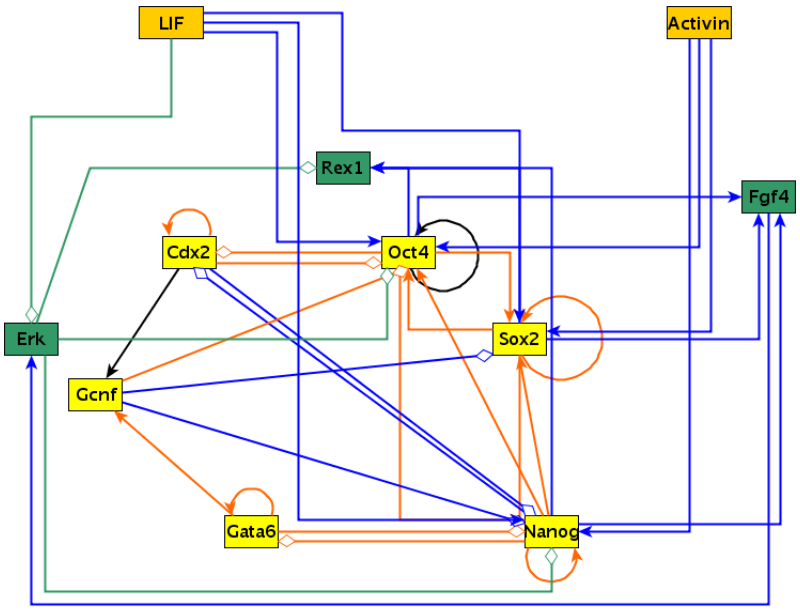


Figure 2.6: The NCC network used for analyzing the role of culture conditions.
The two orange top nodes represent the two culture media, LIF + serum ('LIF') and Activin + FGF2 ('Activin'), while the yellow nodes in the center are the core nodes inherited from the CME and the three green nodes are genes added to model the differentiation from mESCs to mEpiSCs. There are four edge colors, representing the origin of each interaction. Black edges are present in the CME only, orange indicates a common existence in the CME and the PluriNetWork, blue denotes existence in the PluriNetWork only and green stands for the four additional interactions described in the text. Diamond edges denote inhibition, while arrow edges represent activation. Note that the explicit Oct4-Sox2 dimer is removed, as its formation is hidden in the logical functions (a downstream gene is only activated if both Oct4 AND Sox2 are present). This does not change the behavior of the model.

binding sites for LIF/Activin that sterically hinder the binding of Erk?), or expression analysis with microarrays (are Nanog and Oct4 expressed in the presence of Erk only and LIF and Erk together?).
All logical rules for the NCC can be seen in Table 2.7. Note that LIF and Activin have constant values, manually set before the calculation of the state transition graph, so there are no Boolean functions for these two proteins. This manual control of culture media allows us to model the action of the experimentator, removing LIF and driving cells to differentiation. Moreover, it gives us the possibility to combine three hierarchically organized differentiation steps (from zygote to 16-cell stage (A), from ICM to Epi and PE (B) and from Epi to somatic cells (C); Figure 2.7) into one model.

2.4.2 Analysis of the state transition graph

The NCC model can be analyzed with the same methods described above for the CME network. However, the state transition graph can now be calculated under four different conditions (with varying presence of LIF and Activin).
The state transition graphs for various conditions are shown in Figure 2.7 and the attractors

$$
\begin{aligned}
\text{Oct4} &= ((\text{Oct4\&Sox2}) \vee \text{Nanog} \vee \text{LIF} \vee \text{Activin}) \,\&\neg\, (\text{Cdx2} \vee \text{Gcnf} \vee (\text{Erk\&}\neg\text{Activin})) \\
\text{Sox2} &= ((\text{Oct4\&Sox2}) \vee \text{Nanog} \vee \text{LIF} \vee \text{Activin}) \,\&\neg\, (\text{Gcnf}) \\
\text{Nanog} &= ((\text{Oct4\&Sox2}) \vee \text{Nanog} \vee \text{LIF} \vee \text{Activin}) \\
&\quad \&\neg\, (\text{Cdx2} \vee ((\text{Gata6} \vee \text{Gcnf})\&\neg(\text{LIF}))\, (\text{Erk\&}\neg\text{Activin})) \\
\text{Cdx2} &= (\text{Cdx2}) \,\&\neg\, (\text{Oct4} \vee \text{Nanog}) \\
\text{Gata6} &= (\text{Oct4} \vee \text{Gata6}) \,\&\neg\, (\text{Nanog}) \\
\text{Gcnf} &= (\text{Cdx2} \vee \text{Gata6}) \\
\text{Erk} &= (\text{Fgf4}) \,\&\neg\, (\text{LIF}) \\
\text{Fgf4} &= ((\text{Oct4\&Sox2}) \vee \text{Nanog}) \\
\text{Rex1} &= ((\text{Oct4\&Sox2}) \vee \text{Nanog}) \,\&\neg\, (\text{Erk})
\end{aligned}
$$

Table 2.7: Logical functions **F** for the NCC network in Figure 2.6. All rules follow the 'one inhibitor suffices' rational, with the exception of Nanog, which Gata6 and Gcnf can only inhibit if neither LIF nor Activin are present. There are no functions for LIF and Activin as their values remain constant after manually setting them. $\vee$ is a logical OR, & is a logical AND and $\neg$ is a logical NOT.

of these are listed in Table 2.8. In the case of no culture medium being present (-LIF/-Activin) and with LIF only (+LIF/-Activin), a zygote is used as starting state. If LIF is not present but Activin (-LIF/+Activin), the ICM/ESC state is the starting state as Activin is used for either maintaining mEpiSCs or deriving them from mESCs, not from zygotes. The last condition with both LIF and Activin (+LIF/+Activin) is not analyzed in detail as such a culture is not used in the laboratory. Nonetheless, an inspection of its state transition graph shows that it has the same effect as +LIF/-Activin, suggesting a dominance of LIF over Activin - a fact that could be tested experimentally by adding both cultures at once and counting the number of resultant ESCs and EpiSCs. A possible physiological interpretation of the attractors is provided in Table 2.8. It can be seen that the attractors are essentially the same as in the CME, but now we are able to distinguish between ICM/ESC and Epi by the expression or absence of Rex1, respectively (and vice versa for Erk). Only for one condition (-LIF/-Activin), the all-off or somatic cell state can be reached. If either LIF or Activin are present, the somatic state is not steady, as constitutive expression of Nanog, Oct4 and Sox2 are enforced by either culture medium. The (TE) state is similar to the one in the CME network, an impaired TE state with missing Gata6. The difference between LIF and Activin is the reachability of the ESC state, which is steady only with +LIF/-Activin, and the other way round for the Epi and PE states. This is consistent with the known roles of LIF maintaining mESCs and Activin maintaining mEpiSCs. At first glance it might seem unreasonable to obtain the TE and (TE) state even in the presence of LIF or Activin. However, this is no inconsistency of our network as LIF is usually applied to ESCs or ICM cells, and not to zygotes. It would be interesting nonetheless to study the resulting cell types if LIF is applied to zygotes *in vitro*. The attractors of the network under +LIF/-Activin conditions with the ICM/ESC as the starting state do not contain any of the TE-like states. Similar reasoning is valid for Activin: the attractors of the network under -LIF/+Activin conditions with Epi as starting state do not contain the PE state. Note that the ESC attractor does not entirely match with the ICM state given in Table 2.1, as both Gata6 and Gcnf are absent in the

Boolean 'ESC' attractor state. However, the ICM state would not be steady in this network as the inhibition between Gata6 and Nanog would force the ICM state to take one of two routes, either to Epi (with Nanog) or to PE (with Gata6). So we assume that the presence of LIF does not allow suppression of Nanog (cf. also the change in the Nanog rule) and therefore accept the absence of Gata6 and Gcnf in the ESC state. In the following we assume the ESC attractor state to match with the ICM state from the expression table above.
Note that in the -LIF/-Activin case only those states are attractors that are not intermediate on the way to other other cell types (which is the case for ICM and Epi).

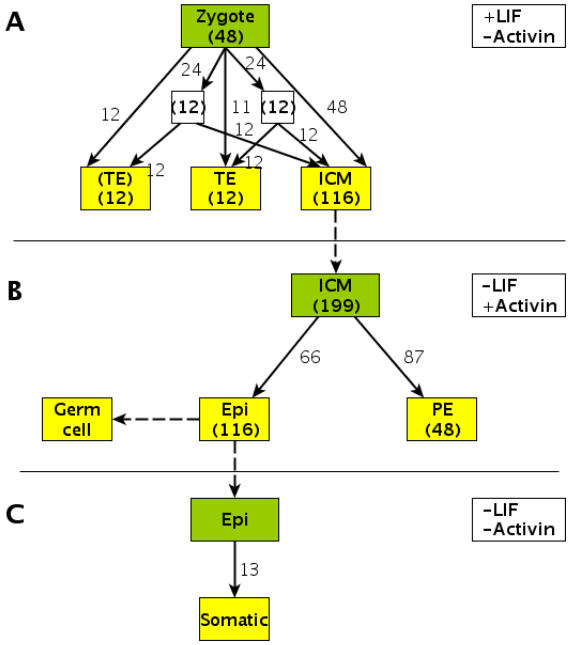


Figure 2.7: The compressed and combined state transition graphs for three different network conditions. The cultures used in each part (divided by the horizontal lines) are written in the top right of each part. A is the state transition graph for a zygotic starting state with LIF on and Activin off. B is the state transition graph for an ICM initial state with LIF off and Activin on. C is the state transition graph for no culture conditions and an Epi starting state. Numbers in nodes and on edges are used as in Figure 2.4. Yellow nodes are attractors, while green nodes represent initial states. The dashed vertical lines across the horizontal lines do not represent a developmental event but rather a re-usage of a former attractor as an initial state under different conditions. Crossing the line between A and B is equal to switching culture conditions from LIF to Activin, while the line between B and C is crossed when Activin is removed from Epiblast cells. Germ cells are derived from Epi cells, but this is not analyzed in detail here.

Three state transition graphs are analyzed: (A) for a zygotic starting state with LIF on and Activin off (+LIF/-Activin), (B) for an ICM initial state with LIF off and Activin on (-LIF/+Activin), and (C) for an Epi initial state with both culture media off (-LIF/-Activin). A picture of each compressed state transition graph is shown in Figure 2.7 (see Figure 2.4 for an explanation of the compression). The hierarchy of parts A and B perfectly matches with the upper part of the development tree in Figure 2.2, if the (TE) and TE states are combined into one TE state. Crossing the upper horizontal line (between A and B) is equal

	LIF and Activin presence										
Gene	-LIF/-Activin					+LIF/-Activin			-LIF/+Activin		
Oct4	1	0	0	0	0	0	1	0	1	1	0
Sox2	1	0	0	0	0	0	1	0	1	1	0
Nanog	1	0	0	0	0	0	1	0	1	1	0
Cdx2	1	1	0	1	0	1	0	1	0	0	0
Gata6	1	1	1	0	0	1	0	0	1	0	1
Gcnf	0	1	1	1	0	1	0	1	1	0	1
Erk	0	0	0	0	0	0	0	0	0	1	0
Fgf4	0	0	0	0	0	0	1	0	1	1	0
Rex1	0	0	0	0	0	0	1	0	1	0	0
LIF	0	0	0	0	0	1	1	1	0	0	0
Activin	0	0	0	0	0	0	0	0	1	1	1
	[Zygote]	TE	PE	(TE)	somatic	TE	ICM	(TE)	[ICM]	Epi	PE

Table 2.8: Starting states (in square brackets) and attractors of the NCC model for different combinations of LIF and Activin. In the left part (-LIF/-Activin) no culture medium is present and a zygote is used as the initial state. In the middle part (+LIF/-Activin) only LIF is present and a zygotic starting state is used, too; this part corresponds to part A in Figure 2.7. In the right part (-LIF/+Activin), only Activin is present and the starting state is an ICM cell; this part corresponds to part B in Figure 2.7. A possible physiological interpretation of each state is provided in the last row.

to switching culture conditions from mESC maintaining (+LIF/-Activin) to Epi maintaining (-LIF/+Activin). One can consider this line as an artificial stop in development, which actually corresponds to the murine diapause, a transient (LIF-induced) developmental stop [158]. The lower horizontal line (between B and C) is traversed by removing the Activin medium from stage B. Note that the differentiation of an Epi into a germ cell cannot be made visible in our network as there is no gene that switches off Fgf4 and Erk.
Interestingly, an Epi cell can be reprogrammed into an ESC by changing culture conditions from Activin to LIF (data not shown, see also [64]).
A betweenness calculation (as performed above for the CME network) is less informative in this case as obviously the ICM and Epi attractors that are reused as initial states have the highest betweenness in the compressed state transition graph.

2.4.3 Stability

The stability of the network and the attractors is calculated as above for the CME network. The local attractor stability, derived from single or double bit flips, is listed in Table 2.9. For calculation details, please refer to section 2.3. Again, the pluripotent states (ESC or Epi) are more stable than the others (0.95 *vs.* 0.73 and 0.91 *vs.* 0.68), which is consistent with the idea of a pluripotent 'ground state' as proposed in [123, 155]. As above, the (TE) and TE state have equal stability, so further indicating the similarity of these two states. The somatic state is rather stable against perturbations, which might be due to the lack of a positive marker for this state.
The global stability is assessed with the individual knock-out of each protein and each gene.

No culture media are used (-LIF/-Activin) for this analysis. To measure the differences after a network disruption the attractors are compared with the original NCC attractors. For each knock-out the number of missing and new attractors is calculated. All disruptions that caused the attractors to change are listed in Tables 2.10 and 2.11 for genes and interactions, respectively. The last column contains a comment and literature references for each knock-out. The results are similar to the ones above in the CME network; for details please refer to the tables.
Additionally, the dependence of the attractors on the starting state is examined. To assess this the state transition graphs are calculated for slightly modified zygotic starting states (one bit flipped in turn) with no culture media and the resulting attractors examined. The results are listed in Table 2.12 in a similar fashion as above with missing and new attractors. It can be seen that the perturbation results are similar to the gene knock-out results, as the lack of Nanog impairs the formation of three tissues, and a missing Cdx2 does not allow for TE formation.

Culture	Attractor	I_1	I_2
+LIF/-Activin	ESC	0.95	0.91
	TE	0.73	0.53
	(TE)	0.73	0.52
-LIF/+Activin	Epi	0.91	0.81
	PE	0.68	0.47
-LIF/-Activin	Somatic	0.82	0.69

Table 2.9: Local attractor stability of the NCC network (Figure 2.6). The higher I_b is, the more likely the attractor will return to itself after a perturbation of $b = 1$ or $b = 2$ bits, respectively. The first column denotes the culture conditions and starting state.

2.4.4 Regulatory motifs in the network

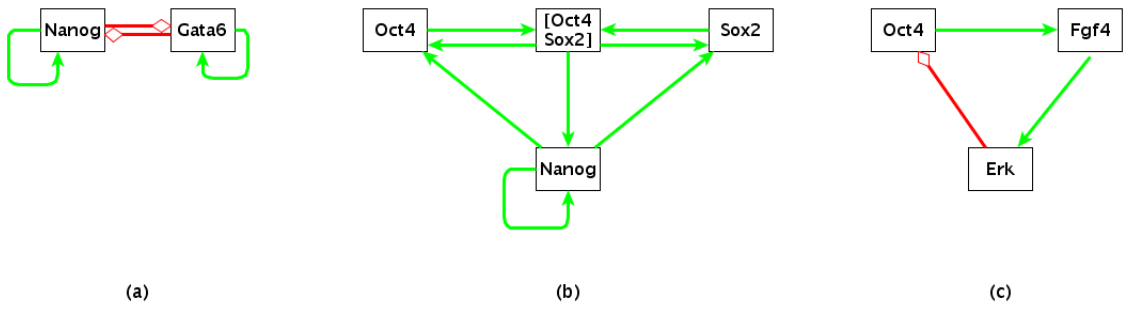


Figure 2.8: Three different network motifs occurring in the CME or NCC network. **(a)** Mutual inhibition with auto-activation, mostly responsible for tilting the balance into either one of two possible lineages. **(b)** Positive feedback loop between three genes, fostering their own expression and thus stabilizing the states which require their expression. **(c)** Suppressive feedback loop, where one gene activates two others in chain, eventually causing its own downregulation.

KO Gene	M	N	Comment
Oct4	-	1	A state with Sox2, Erk and Fgf4 only occurs; this could be an impaired Epi state, underlining the importance of Oct4 in pluripotency [23]
Sox2	-	-	(no changes); loss possibly compensated by Nanog and Oct4
Nanog	2	-	(TE) and somatic state are missing; consistent with the high importance of Nanog in development [26, 27, 114, 155, 16, 126]; the missing (TE) state could be explained by a required intermediate state with Nanog
Cdx2	2	-	(TE) and TE are missing; consistent with Cdx2 requirement for TE [128]
Gata6	2	-	PE and TE missing; consistent with Gata6 expression in both tissues [22, 62]
Gcnf	-	-	(no changes); consistent with the enforcing, but not decisive role of Gcnf
Erk	0	1	An ESC state occurs; consistent with Erk necessary to switch off pluripotency [94]
Fgf4	0	1	An ESC state occurs; consistent with Fgf4 requirement for differentiation [18, 96]
Rex1	-	-	(no changes); Rex1 does not seem to play a role in the establishment of the four original attractor states

Table 2.10: Global stability as knock-out analysis of genes in the NCC network. For each change the state transition graph with a zygotic starting state is calculated, and the attractors are compared with the original NCC attractors with LIF and Activin off. The first column shows the knocked-out gene, column 2 lists the number of missing attractors (M), and column 3 the number of new attractors (N). In the comment column the changes are listed in detail for each disruption.

It has been shown in several publications that most, if not all, biological networks consist of few basic building blocks, so-called motifs, which play decisive roles in the network dynamics [4]. In the CME and NCC network (Figure 2.6), several regulatory network motifs are present. There are mutual inhibitions between pairs of genes which auto-activate themselves (see Figure 2.8(a)) - a motif which can create a distinction between two different trajectories (cell lineages) in the state transition graph. Namely, these are Oct4 and Cdx2 (which distinguishes between the ICM or TE lineages), Nanog and Gata6 (which distinguishes between the Epi and PE lineages), and Nanog and Cdx2 (which does not directly correspond to a lineage decision, but obviously fosters the distinction between the ICM/Epi and the TE lineage). The triad of Oct4, Sox2 and Nanog comprises a self-reinforcing loop (see Figure 2.8(b)), which is responsible for the stability of the pluripotent states. There is also an incoherent feed-forward loop (in the CME only), with Oct4 activating Gata6 and Nanog, but Gata6 also inhibiting Nanog, creating the suppression of Nanog at high Oct4 levels (not in our Boolean network). Of importance for the downregulation of the pluripotency triad are the suppressive feedback loops (see Figure 2.8(c)), e.g. Oct4 activating Gata6, which activates Gcnf and this gene in turn suppresses Oct4, or the activation of Erk via Fgf4 by Nanog and Oct4, eventually causing their own downregulation.
There are further motifs in the NCC network that are not described in detail, but it is

KO Interaction	M	N	Comment
Sox2 → Oct4	-	1	A new state with Oct4 as the only active gene arises; so the dimerization of Oct4 and Sox2 seems to play an important regulatory function [32]
Erk $\overset{-}{\rightarrow}$ Oct4	-	1	An Epi-like state without Nanog is new; Erk might thus play a role in enforcing the decision to leave pluripotency
Oct4 → Sox2	-	1	A new state with Sox2 as the only active gene arises; so the dimerization of Oct4 and Sox2 seems to play an important regulatory function [32]
Erk $\overset{-}{\rightarrow}$ Nanog	-	1	An Epi-like state without Oct4 is new; Erk might thus play a role in enforcing the decision to leave pluripotency
Cdx2 → Cdx2	2	-	Both the TE and (TE) states are missing; consistent with the high relevance of Cdx2 in the formation of TE [128]
Nanog $\overset{-}{\rightarrow}$ Gata6	2	-	The two missing states are somatic and (TE); this seems to indicate a requirement of Gata6 suppression for the formation of these tissues
Gata6 → Gata6	2	-	Missing are PE and TE; consistent with the expression of Gata6 in both tissues [22, 62]
Cdx2 → Gcnf	1	1	The (TE) state has no Gcnf expressed
Gata6 → Gcnf	1	1	The PE state has no Gcnf expressed
Fgf4 → Erk	-	1	An ESC state arises; consistent with the repression of this state by the Fgf4/Erk chain [94]

Table 2.11: Global stability as knock-out analysis of interactions in the NCC network. For each change the state transition graph with a zygotic starting state is calculated, and the attractors are compared with the original NCC attractors with LIF and Activin off. The first column shows the knocked-out interaction, column 2 lists the number of missing attractors (M), and column 3 the number of new attractors (N). In the comment column the changes are listed in detail for each knock-out. Only those disruptions are listed that actually cause the attractor set to change. A → B denotes activation of B by A, whereas A $\overset{-}{\rightarrow}$ B denotes an inhibitory function of A on B.

already obvious from the few listed ones that the dynamic patterns of the complete network are only a combination of the simple dynamics of the motifs. Thus, the full Boolean network could actually be reduced to its motif components [14], allowing for the assessment of more complicated networks.

2.5 Discussion

Boolean modeling in general can be used to gain insights into the dynamic capabilities of GRNs on a system level. In this chapter we have constructed two Boolean networks which try to capture the early development of a fertilized murine egg cell, both with physiological cell types and the timing and hierarchy of the involved lineage decisions.

Gene	M	N	Comment
Nanog	2	-	(TE) and somatic state are missing; consistent with the high importance of Nanog in development [26, 27, 114, 155, 16, 126]; the missing (TE) state could be explained by a required intermediate state with Nanog
Cdx2	2	-	(TE) and TE are missing; consistent with requirement of Cdx2 for TE [128]
Gata6	2	-	PE and TE are missing; consistent with the expression of Gata6 in these tissues; this indicates that Gata6 must either be activated intrinsically (e.g. by Oct4, see 2.3) or be present in the zygote

Table 2.12: Dependence of the attractors on the starting state. Each bit in the zygotic starting state (no culture conditions applied, -LIF/-Activin) is flipped in turn, and the change in the attractor set of the state transition graph in comparison with the original NCC attractors is reported here. The second column lists the number of missing attractors (M), and column 3 the number of new attractors (N). In the comment column the changes are listed in detail for each flip. Only those flips are listed that actually cause the attractor set to change.

2.5.1 Result summary

The state transition graphs of the networks contain attractors and hierarchies that resemble the true zygotic development astonishingly well. Although no prior information about this hierarchy is used when building the network, the topology of the involved GRN and the logical rules convey it. It has been shown that culture conditions can be modeled appropriately with a Boolean network. The roles of essential genes during embryo maturation have been analyzed and the findings are conform with literature. The first network has been compared to an ODE modeling approach of murine zygotic development, and it has been shown that the essential dynamic properties are sufficiently covered by a Boolean network, too, but the necessity for any parameters is avoided. This requirement for qualitative data only is one of the great advantages of a Boolean network. Given the abundant lack of quantity in mESC studies, mostly due to the complexity of the researched system, a qualitative model thus seems a natural choice. Furthermore, the stability of the networks against perturbations has been assessed, confirming results of earlier analyses.

From our network analysis we conclude two hypotheses that require further investigation in the laboratory. The first one states that the presence of LIF does not allow the suppression of Nanog by Gata6 or Gcnf, and it is motivated by the disequality of the ICM/ESC attractor state and the expression values curated from literature. In our models the ICM state has neither Gata6 nor Gcnf expressed, although studies suggest that. One possible explanation is that the ICM state would not be stable if Nanog and Gata6 were expressed at the same time, as they exert a mutual inhibition. However, it could be possible that Gata6 is maintained at a low basal level that is not 'visible' in our Boolean network. A recent study showed that Nanog and Gata6 are expressed in a salt-and-pepper fashion in the ICM, with some cells expressing Nanog and others Gata6, but no cells expressing them simultaneously [28, 157]. Following this point of view, we only see the Nanog-expressing ICM cells in our network. A different explanation could be the direct suppression of Gata6 and Gcnf by LIF.
Our second hypothesis is similar, incorporated in the Nanog and Oct4 logical functions. Both genes are known to be inhibited by Erk, and we suggest that this inhibition is only possible

if Activin is not present. This reasoning results from the knowledge that Activin maintains mEpiSCs for indefinite time *in vitro*. To check these two proposals in the laboratory a promoter inspection of Nanog, Oct4 and Gata6 together with a microarray gene expression analysis could create valuable insights.
Note that the 'one inhibitor suffices' logical rational is not directly motivated by experimental findings but is rather a general idea about the regulatory nature of transcription factors. Thus proposing a different function for developmentally highly important genes as Nanog and Oct4 is not contradictory by itself, but rather shows a glimpse of the bandwidth of possible regulations in a cell.

The attractors of both Boolean networks match with biological cell types, identified as PE, TE, ESC, Epi and somatic cells by their expression values. The only imperfect match is the (TE) state which lacks expression of Gata6. As explained above, an activator of Gata6 in the TE tissue would cure this problem, but to our knowledge there is none within the set of genes that are included in the network. One could interpret the (TE) state as an intermediate state on the path to TE, which transiently has switched off Gata6; this remains to be tested biologically. Evidenced by the close grouping of the (TE) state with the TE state in the state transition graphs and the equal stability against perturbations, the difference between these states is not a major problem.
All logical rules used in the networks follow the rational of 'one inhibitor suffices', with the exceptions of the heterodimer Oct4-Sox2 and Nanog and Oct4 (in the NCC network only). Although this rule might seem too simplistic, it is sufficient to capture the essentials of mESC development. Interestingly, there is increasing evidence that many developmentally important genes in mESCs have a 'poised' state. They bear both activating and inhibiting histone marks [6, 10, 127] or their promoter is occupied by activators and inhibitors at the same time [126]. One explanation for this phenomenon is to have the genes ready for fast transcription if a development switch toggles [126] and the (single) inhibitor is removed. Nonetheless, one could try to infer the logical rules from literature, either by large-scale promoter occupancy studies similar to [85, 15, 101], or by the nature of regulation shown in microarray and knock-out studies like [109].
Note that the missing notion of time makes it hard to interpret intermediate states or path lengths in the transition graph (e.g. in the betweenness calculation). These can be used as a primer for intuition, but do not necessarily have any biological meaning.

The benchmarks for our networks (the development tree in Figure 2.2 and the expression values in Table 2.1) are curated from different data sets and the previous work of many groups. On the one hand, this has been necessary as to our knowledge there is no study that provides this table as one, but on the other hand, this requires to take special care of the comparability of the sources. For our purposes we have considered two expression studies as comparable, if they are performed for the same cell type (e.g. mESC), and the fold change between expression or absence is at least two. The equality of gene, mRNA and protein is a necessary simplification, as there has been few proteomics data for mESCs up to now (see e.g. [8]). We argue that this assumed equality is adequate for a qualitative network.

The moderate network size (11 players for the NCC network) is a result of the following two considerations. On the one hand, the size of the state space rises exponentially with the network size, and also the number of artificial attractors rises with the network size [89]. Thus keeping the network as small as possible as long as it explains the data is a reasonable design principle, conform with Occam's razor. On the other hand, for each of the players there is abundant literature about their roles in pluripotency and development, which

is a clear motivation to include them. Additionally, the knock-out analysis of both genes and interactions yielded that most of them are required, as disruption changed the network behavior. Considering that the knock-outs are single disruptions, unable to capture the effect of simultaneous failures (e.g. synthetic lethals), it is reasonable to assume that all included genes and interactions are of importance.

It has to be mentioned that some of the knowledge found in the literature is not considered in our networks. For example, Nanog actually requires homodimerization before acting as a TF [119]. This is not included since protein numbers are not part of a Boolean network, so the dynamics of the system would not change. The lacking inclusion of the Oct4 dose dependence [125, 131] has been discussed above. However, dose dependence does not necessarily imply that the level of a protein is exactly the same in all cells, but rather that the average is within a strict range, while the amount in individual cells can differ [139]. Thus we argue that the simplification to two values in our networks is reasonable at least for a subpopulation of mESCs. For the same reasons the auto-inhibition of Oct4 mentioned in the PluriNetWork is not considered, as this suppression is possibly necessary to maintain Oct4 at a precise level in pluripotency [131] or to reduce noise effects [139]. Nonetheless, an incorporation of more than two (or continuous) Oct4 levels is desirable and could be treated with ODEs (if kinetic rates are available) or discrete models (as proposed in [186]). The expression of Rex1 is used as discrimination between mESCs (on) and murine EpiSCs (off), but only *in vitro* [18], and two recent studies indicate that this view might be too simple [64, 174]. Nonetheless, in our network Rex1 is used as a mESC marker. A recent publication states that an overexpression of Sox2 inhibits Fgf4 [13], indicating a necessity for intermediate levels not only for Oct4, but also for Sox2. This is, as discussed for the Oct4 case, not considered in our networks.
From the heterogeneity and dose dependence of various genes in the network we can also conclude that time-resolved data of single cells are of utmost importance as only these can help to gain insight into the noisy nature of the genes.

Summing up, our Boolean networks reproduce the established biological patterns despite inherent drawbacks of the modeling approach and minor inconsistencies. This underlines the importance of considering a simple approach first before delving into more detailed models.

2.5.2 Outlook

Of course, the embryonic development is not finished when the somatic state is reached. In contrast, the true differentiation into the different cell types of the body merely begins at this point, and it is a thrilling challenge to model this part of development, coarsely shown in Figure 1.2, with a Boolean network. The opposite direction is the reprogramming of a terminally differentiated cell back to pluripotency, as outlined in [167] and numerous subsequent studies (nicely summarized and reviewed in [68, 97]). It is desirable to know whether such a reprogramming process is equal to a reverse walk along the developmental trajectories, or rather a way completely off the beaten paths. To study this in a Boolean network one would have to generate mouse embryonic fibroblasts (MEFs) along their default developmental path, then add the reprogramming factors (Oct4, Sox2, Klf4, c-Myc) and inspect the pathway of the following events.
Another possible extension of the networks proposed here is the inclusion of epigenetic modifiers like Polycomb group proteins (PcG), responsible for the accessibility of the genome in ESCs [6, 126, 110], or microRNAs, which are known to play essential roles both in the maintenance of pluripotency [56, 70, 180] and in development [35, 75, 133].

An approach currently studied in [103] (unpublished work) is to supply a network selection algorithm with a number of genes and a coarse interaction structure, along with an expression table with missing values similar to Table 2.1. The algorithm then could perform a search for those networks/interaction structures that are able to reproduce the provided expression values as attractors, and additionally fill the missing values in the expression table with reasonable entries. If one value stays constant over all networks, one could deduce a testable hypothesis about the state of this gene. This network sampling might also be a measure for the uniqueness (or privilege of interpretation) of the provided network structure.
An interesting question is the applicability of our networks for other species, especially humans. All the networks and statements above have been made for the mouse model and do not necessarily apply for human embryonic development. Currently the overwhelming amount of data is available for mouse (e.g. the PluriNetWork is for mouse only). Nonetheless, the human developmental pathway [140] is quite similar to mouse and most of the genes used in our networks are also known to be essential for human embryonic stem cells (hESCs) or development as well (Nanog, Oct4, Sox2: [7, 15]; Other genes: [21, 40]), though the details are often critically different [58, 147, 140]. Especially the culture conditions act differently, as LIF is known to fail in maintenance of hESCs [39], while Activin culture can maintain hESCs [77]. This suggests that hESCs are closer to mEpiSCs [171]. Altogether, the core nodes and interactions of our networks should apply for human development, too, but some essential differences currently limit the applicability. An extensive literature search to check all interactions could change this.

In the next section, the opposite of a network extension is analyzed. One gene that is essential in pluripotency is picked from the network and studied in more molecular detail: Nanog, the 'land of the ever young' [26].

Chapter 3

Nanog dynamics in single embryonic stem cells

3.1 Nanog and its role in pluripotency

In 2003, Chambers et al. finally identified a protein that had already been anticipated to play an essential role in sustaining the pluripotency of mESCs [27]. The authors performed a large-scale functional screening approach, while another group executed a combined bioinformatics and experimental approach [114] and found the same transcription factor. This protein was named Nanog by Chambers, after the mythological Celtic land of the ever young because its overexpression enables mESCs to proliferate indefinitely, *i.e.* to stay 'forever young'. It is a homeodomain protein located on chromosome 6 (in mouse) which comprises 305 amino acids and is conserved in human, chimpanzee, dog, cow, rat and mouse.[1] It builds homodimers [119] and functions as a transcription factor regulating key eukaryotic development processes. Nanog expression had been considered a requirement for pluripotency [26, 114], but in 2007 the discoverers of Nanog themselves brought this role into question [27]. Chambers et al. found that the levels of Nanog are actually heterogeneous in mESCs, and that even cells with virtually no Nanog expression still have the potential to form colonies and contribute to all three germ lines (mesoderm, endoderm and ectoderm) - a unique feature of pluripotency. Along with Nanog, heterogeneity in expression was also reported for other pluripotency marker proteins, namely Rex1 [174] and Stella [66]. Thus the former consideration of pluripotency as a state with stable gene expression values went awry. Now pluripotency is rather assumed to be a highly dynamic state [61, 116, 160] with heterogeneous expression values of certain genes. Proposed explanations are the robustness against small perturbations (by not requiring stiffly regulated protein levels) or quick response times to external stimuli (by fostering one expression regime over the others) [61, 160].
After the initial description of Nanog heterogeneity subsequent studies [60, 78] tried to unravel the mechanism behind these fluctuations of Nanog. The authors in [60] propose two different computational models that both explain the observed bimodality of Nanog intensities and also suggest experimental strategies to distinguish between the two of them. In [78] another model is suggested and analyzed, both theoretically and experimentally. The lacking conformity of these three models with our experimental observations is discussed below and a new model is proposed.

[1] Information extracted from PubMed, with Gene ID 71950 and Protein ID AAI44997.1.

Today the molecular mechanism behind the heterogeneous expression of Nanog is still unknown, and in this chapter we try to shed light on it by taking advantage of novel and unique high-quality observations of mESCs on population and single cell level. A workflow of the computational analysis can be seen in Figure 3.1. All laboratory experiments were conducted by Adam Filipczyk (AF) at the Institute for Stem Cell Research (ISF) at the Helmholtz-Zentrum München and have not been published yet. The close and fruitful collaboration of the Computational Modeling in Biology (CMB) group at the same center with the ISF group is of vital role for this thesis and also for current and future research.

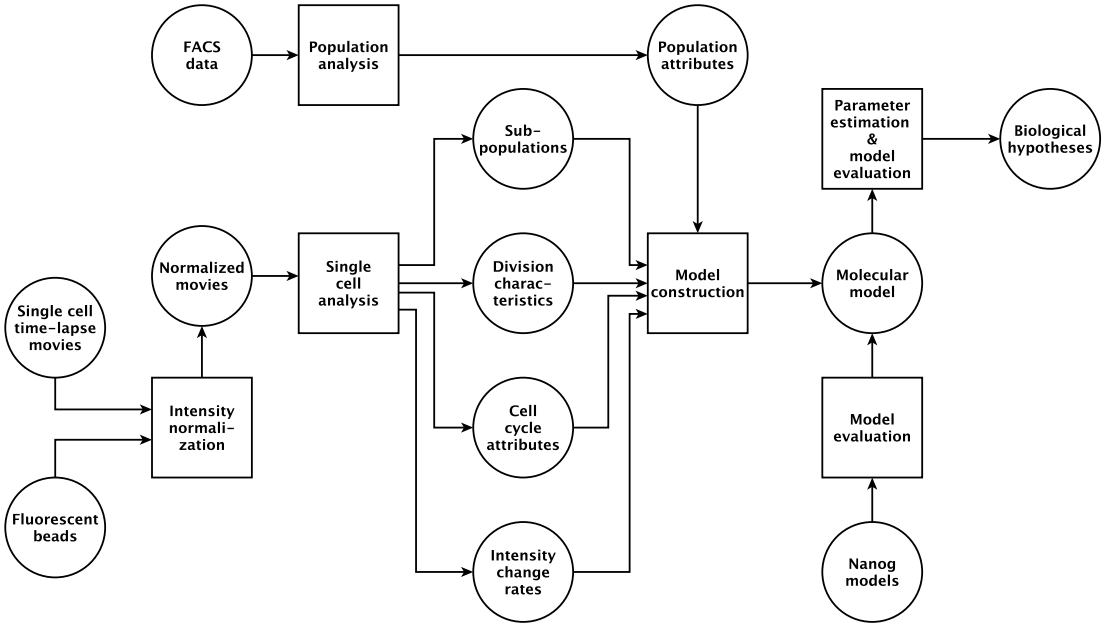


Figure 3.1: The Nanog fluorescence analysis pipeline. The workflow is modeled as a Petri Net, a bipartite graph connecting states (data or results, in circles) with transitions (actions, in squares) [136].

3.2 Data types and sources

3.2.1 ES cell lines

Chambers et al. [27] and a subsequent study [78] used a transcriptional reporter mESC line to quantify Nanog expression in a FACS (fluorescence activated cell sorting) analysis. This cell line has one allele of the Nanog gene removed and replaced by an eGFP (enhanced green fluorescent protein) gene sequence, allowing for quantification of the transcription of Nanog by measuring the fluorescence intensity of eGFP.
In this study, data from a NanogeGFP fusion ESC line are analyzed. This cell line has one

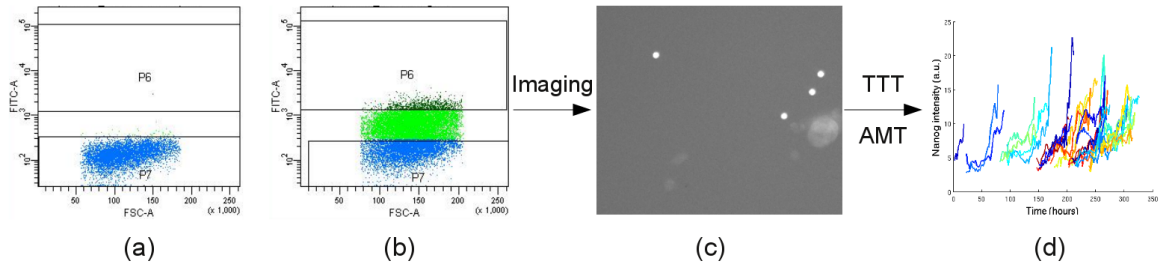


Figure 3.2: Schematic depiction of the data generation in the laboratory. **(a)** E14 control cells FACS plot. P7 is the negative gate that contains all control cells, P6 is the gate defined as the brightest five percent of Nanog positive cells (not necessary for the control cells, see (b)). The x-axis is the forward scatter (FSC), which sorts cells according to their size, the y-axis is the eGFP intensity. **(b)** ESC fusion line FACS plot. Axes are labeled as in (a). Now five percent of the cells are in gate P6 (actually the definition of the gate). **(c)** Section of a fluorescent image of one ESC time-lapse movie. The bright balls are the fluorescent beads utilized to calibrate the microscope and to compare the fluorescence intensity between different positions. The other shining spots are ESCs which express Nanog. **(d)** Quantification of a cell genealogy. Each curve displays the intensity of a single cell. The x-axis denotes experiment time in hours, the y-axis denotes the Nanog intensity in arbitrary units. The intensity of each cell is rising during the growth phase of the cell cycle and is approximately halved during division.

Nanog allele replaced by a sequence comprised of the Nanog sequence and the eGFP sequence in the 3'-end, so that the resulting protein is a fusion of Nanog and eGFP (at the Nanog-C-terminus). This construction allows for a more authentic quantification of Nanog protein abundance. In the following ESC always refers to this fusion cell line if not explicitly marked. All cell lines were cultured under +LIF (with fetal calf serum) conditions to maintain the pluripotent ESC state (see chapter 2 for details about LIF).

3.2.2 Fluorescence activated cell sorting (FACS)

Both ESC lines described above (reporter and fusion line) were subjected to FACS analysis to quantify the intensity of eGFP and thus the expression of Nanog in a population of individual cells at single timepoints. In a first step the intensity of unmodified control cells (E14 cells, see e.g. [25, 192]) was measured with the same laser wavelength used for eGFP activation (488 nm). These cells define the background intensity that is present by default and marks the cutoff between Nanog positive (expressing) and negative (not expressing) cells in the later analyses (see Figure 3.2). The intensity region of these control cells is called the negative gate, and all cells within this gate are considered negative. The second step was to quantify the intensity of the modified cells (either reporter or fusion line) and to divide them into positive and negative cells. With the reporter ESCs a bimodal Nanog intensity distribution with a large peak (about 70% of the cells) in the positive gate and a small peak (about 30%) in the negative gate was observed (see the figures in [27, 78]). A bimodal distribution is termed "macroheterogeneous" [71]. With the fusion ESCs a skewed distribution was observed, but not a bimodal one - there is only a peak in the positive gate, while the negative gate is trailing off to the left without a peak. Such a distribution is called "microheterogeneous" according to [71]. Nonetheless, the numbers within each gate were roughly the same (20-30% in the negative and 80-70% in the positive gate). The discrepancy between the cell lines could be explained by

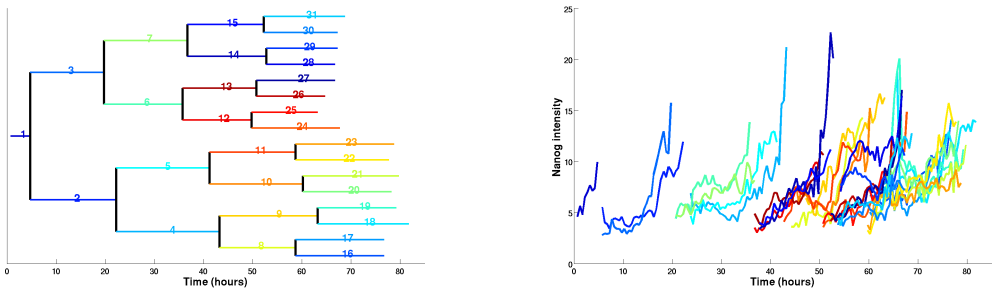
the different half-lifes of eGFP (5.5 hours [38]) and the NanogeGFP fusion (about 3.6 hours, unpublished data). There could be a bias towards higher values in the reporter line data, since eGFP can still be present even though Nanog has already decayed, which produces the negative peak. The fusion line does not suffer from the problem of differing half-lifes, as both the Nanog and NanogeGFP half-lifes are equal (Western blot analysis by AF, unpublished data). Additionally, the fusion line is sensitive to post-translational modifications (PTM) of Nanog and possible regulations on the Nanog mRNA level, resulting in a high correlation between eGFP intensity and true Nanog abundance. Note that the measured intensities represent only half of the real amount of Nanog since only one allele is replaced by the fusion protein. This, however, does not change the qualitative and quantitative conclusions if equal transcription from both loci is assumed. A more detailed method description of FACS can be found in Appendix 5.5.

3.2.3 Single cell time-lapse movies

The fusion ESCs were subjected to long-term (about 7 days) time-lapse microscopy resulting in temporally and spatially highly resolved time-lapse movies. The cells plated on the wells were unsorted, a random sample of the whole population of ESCs with diverse Nanog expression levels. Pictures with fluorescence activation of eGFP were taken every 30 minutes, and the resolution was about 1μm per pixel. From these movies, the cell genealogies were reconstructed for four generations (about 66.2 $\pm$ 19.8 hours) with Timm's Tracking Tool (TTT, see appendix 5.5 for details). For each cell and picture the normalized fluorescence intensity was quantified with Aided Manual Tracking (AMT, see appendix 5.6 for details). The intensity is thereby defined as the sum of pixel intensities within the cell nucleus (defined by a fluorescent nuclear membrane) and the image background subtracted. As we are interested in total Nanog abundance, different cell sizes do not have an effect due to the linear correlation of pixel intensity with Nanog monomer counts (see below). Altogether 31 trees with 873 cells (on average 28.16 per tree) were tracked and quantified, yielding 28,077 individual data points (on average 32.16 per cell). Please refer to Figure 3.2 for a depiction of the data generation process, and to Table 3.1 for the number of trees in individual movie positions. Figure 3.3 shows the genealogy and detailed time course of one example tree.

3.2.4 Bead normalization

In order to account for different illumination strengths of the microscope in different experiments and movie positions (subsections of a well) a normalization is performed. The AMT tool (see Appendix 5.6) already normalizes within one position by correcting for the intensity differences between picture center (brighter) and margin (darker) and temporal differences, as the pictures get darker over time due to photobleaching. However, a normalization across different movie positions is not available, so fluorescent beads were analyzed for each position. These beads are tiny fluorescent balls, slightly smaller than cells, and designed to have equal size and about the same intensity each. Thus if the beads are brighter in one position than in another, we know that the first one is brighter as a whole than the latter one and the cell intensities need to be corrected by a bead intensity factor. It is set to 1.0 in an arbitrary position P_0 and for all other positions P_x the ratio of the bead intensities $R_I = \frac{i(P_x)}{i(P_0)}$ (with $i(P_x)$ the intensity in position x) is the relative intensity of this position. The correction



(a) (b)

Figure 3.3: **(a)** One cell genealogy. Each vertical line corresponds to a division event where the mother cell splits up into two daughter cells. The horizontal lines represent the life times of the cells, with the x-axis in hours. The number of each cell is written above its life line. **(b)** Time course of Nanog expression in this tree. The x-axis denotes movie time in hours, the y-axis the Nanog intensity (linear scale). Each line represents the trajectory of one cell, with matching colors to (a).

factor by which the intensity of the cells is multiplied is $\frac{1}{R_I}$. The position 100728AF1/042 is chosen as a reference and its relative intensity is set to 1. Table 3.1 summarizes the average bead intensities and the relative intensities for all positions.
The beads are bleaching over time with each fluorescent picture acquired, rapidly within the first hours and slowly over the rest. In order to ensure comparability, beads were quantified at the beginning of each movie, in the middle and at the end. The middle and end beads are used to check whether the image normalization works (they should exhibit comparable trajectories across all positions), while the beads from the beginning of the movies are used to calculate the average bead intensity (which is the mean intensity of the first five measured timepoints, averaged over all beads from this position).

3.2.5 Other data

The decay rate of Nanog was determined independently with Western blot and image analysis, yielding a half-life of approximately 3.6 hours. The absolute Nanog protein abundance was quantified with Western blots, indicating an average number of Nanog monomers per cell of approximately 1 million.[2] The decay rate of Nanog mRNA in mESCs was measured by Sharova et al. [153] and is about 5.8 hours.

3.3 Data analysis

3.3.1 Analysis of the FACS population data

The fusion ESC line shows the Nanog intensity distribution pictured in Figure 3.4 if subjected to FACS analysis. As mentioned above, there is no peak in the left part, possibly due to the more sensitive fusion of Nanog with eGFP instead of a transcriptional reporter.

[2]Note that the observed count depends on the sensitivity of the utilized antibody and is possibly subject to change. Nonetheless, several recent experiments by AF revealed numbers of the same magnitude.

Experiment	Position	Bead intensity (sd)	Relative intensity	Tree count
100205AF1	061	20.66 (2.78)	2.47	2
	066	21.48 (1.77)	2.57	2
	068	20.50 (3.65)	2.45	2
100505AF1	011	13.01 (2.42)	1.55	7
	020	11.55 (1.76)	1.38	2
	022	13.71 (2.23)	1.64	1
100609AF1	001	6.33 (0.47)	0.76	3
	002	19.86 (2.14)	2.37	2
	003	20.34 (1.96)	2.43	3
	004	9.62 (3.20)	1.15	2
	010	23.46 (2.41)	2.80	1
	057	21.94 (1.56)	2.62	1
	066	20.06 (2.34)	2.40	2
100728AF1	042	8.37 (1.63)	1	1

Table 3.1: The average fluorescent bead intensities and the relative intensities for each position. The third column is the average intensity (with standard deviation) of the first five measurements of all tracked beads in this position. At least 10 beads are tracked per position, both from the picture center and margins. The fourth column lists the relative intensity, where the last position 100728AF1/042 was arbitrarily set to 1. Additionally, the number of trees in each position is provided in the last column (sum = 31).

In [27] the authors claim that cells in the low Nanog expression (LN) state are able to give rise to cells with high Nanog expression (HN). To prove this they sort their reporter ESC line (with FACS) into LN and HN cells with high purity and culture them for six days under +LIF (with serum) conditions. After this time they see a reestablishment of the initial bimodal Nanog distribution with the large HN peak and the smaller LN peak, indicating the capability of the LN cells to give rise to HN cells and of the HN cells to give rise to LN cells. However, the original distribution of Nanog intensities is not completely restored, as especially the LN sorted cells stay biased. Additionally, the authors imaged single cells in a movie for 44 hours and analyze the transitions from the LN to the HN state. In this movie they observe LN cells that express Nanog after 24 hours. Transitions from HN to LN are not reported from the movies, but their results and one of their figures (1b) suggest it.
Summing up, their results indicate that the Nanog expression levels are not rigid, but cells from either state can give rise to the other.
In the ISF laboratory the fusion ESC line was subjected to a similar analysis. From the original skewed (but not bimodal, see above and Figure 3.4) Nanog distribution the brightest five percent and the dimmest two percent were sorted and cultured for fourteen days. Every second day a FACS analysis of the Nanog intensities was performed and the distribution recorded. This experiment was done twice. The filtering gates were defined as above (Figure 3.2), with LN the complete negative gate (not only the dimmest two percent, but also brighter cells that are all within the range of the E14 control cells) and HN-5 the gate with the brightest five percent of all cells. The data (summarized in Table 3.2) suggest that low sorted cells stay biased even after 14 days, and high cells seem to restore the original intensity distribution better (since LN reaches its original state of 25-30%), although the HN-5 gate

keeps to contain more than five percent. It even seems that the reestablishment gets stalled in the low sorted cells as there is no strong change in the fraction of cells in the LN gate after day 4. Note that two weeks correspond to roughly 24 generations with an average life cycle of 14 to 16 hours (see below).
A similar experiment was performed for the duration of six weeks instead of two weeks, and after this time the original Nanog distribution was completely reestablished (data not shown). This leaves us with the possibility that the bias after two weeks is either real or stems from a problem with the data generation or analysis.
In the following, we assume that a sorted population of cells can give rise to the full Nanog distribution after a certain number of generations, consistent with [27] and our six weeks experiment. The exact time until the full reestablishment of the distribution and the underlying mechanism are, however, not extractable from either FACS analysis.

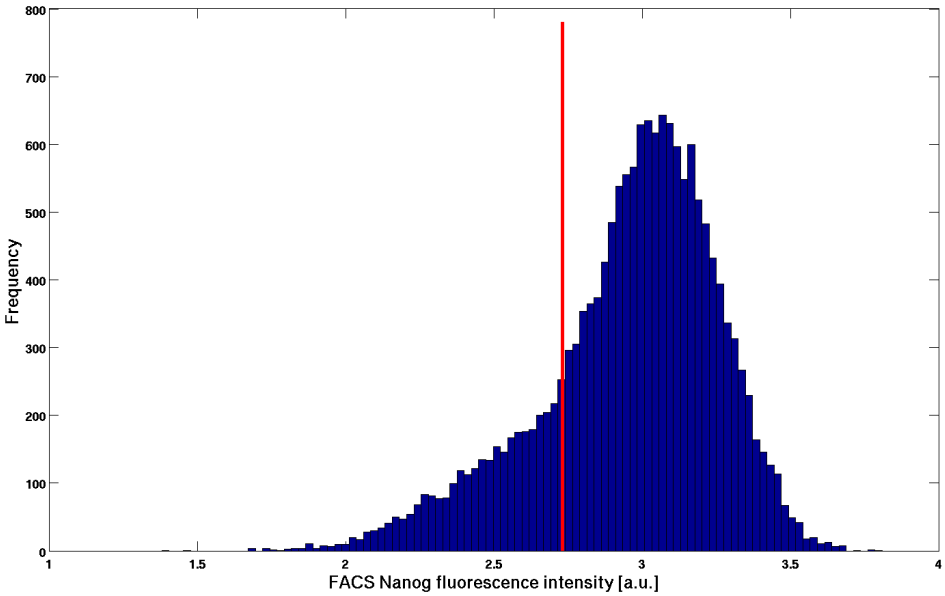


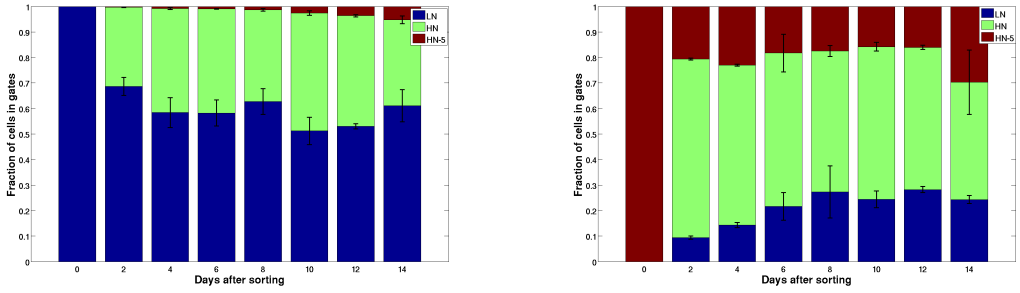
Figure 3.4: FACS Nanog intensity distribution. The x-axis denotes the NanogeGFP intensity (arbitrary units, log-scale), the y-axis denotes the number of cells. The curve is strongly skewed to the left, but shows no bimodality as in other studies [27, 78], possibly due to the different ESC line (see text for details). The red line represents the upper bound of the negative gate. In FACS this is defined by the highest intensity of the E14 control cells. All cells left of the line are considered negative, while all cells on the right are considered positive for Nanog expression.

3.3.2 Nanog expression levels in the time-lapse movies

The FACS data from above summarize the behavior of a population of ESCs, but the question how individual cells function to generate this skewed distribution of the Nanog intensity remains unresolved. In order to gain insights into single cell properties the mESC time-lapse movies are analyzed. The tracking and quantifying has been done by AF in the ISF lab, and in this section the characteristic features of the cells are analyzed.

	LN sorted cells		HN sorted cells	
Day	Cells in HN-5 (sd)	Cells in LN (sd)	Cells in HN-5 (sd)	Cells in LN (sd)
0	0.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)
2	0.01 (0.01)	0.68 (0.04)	0.21 (0.01)	0.10 (0.01)
4	0.01 (0.00)	0.59 (0.06)	0.23 (0.00)	0.15 (0.01)
6	0.01 (0.00)	0.59 (0.05)	0.18 (0.07)	0.22 (0.06)
8	0.02 (0.01)	0.63 (0.05)	0.18 (0.02)	0.28 (0.11)
10	0.03 (0.01)	0.51 (0.06)	0.16 (0.01)	0.25 (0.04)
12	0.04 (0.01)	0.53 (0.01)	0.16 (0.01)	0.28 (0.01)
14	0.05 (0.01)	0.62 (0.06)	*0.30 (0.13)*	0.24 (0.01)

Table 3.2: FACS analysis results of the fusion ESC line. Values are from two experiments, and are given in fractions of cells. HN-5 is the gate of the brightest five percent, LN is the complete negative gate (about 20-30%). The analysis was started with highly pure (99% sorting accuracy, data not shown) cell fractions, and the rows show the data for subsequent days. The HN sorted cells seem to repopulate the entire distribution better as they are less biased after 14 days than the LN cells. A complete repopulation (HN-5 = 0.05 and LN = 0.2-0.3) is not observed during this time, however. The numbers in italic in the HN-5 gate after 14 days of the HN sorted cells are probably a measurement error, indicated by the high standard deviation.



(a) (b)

Figure 3.5: Visualization of Table 3.2. The bars show the fractions of cells in each gate, for prolonged periods after sorting (x-axis). Cells in day 0 have a $> 99\%$ purity. **(a)** Low sorted cells **(b)** High sorted cells

First of all it needs to be clarified whether the imaging conditions are adequate and do not alter the cell behavior with respect to the FACS analysis. The cells are viable and stay pluripotent over the whole period of the analysis, which is proven by tetraploid aggregation (data not shown, experiments by AF). The Nanog intensity histogram, generated from all cells in the movies, is shown in Figure 3.7(a). It seems comparable to the Nanog distribution obtained from the FACS analysis, shown in the comparative density plot in Figure 3.6. The three lines are kernel density estimations and correspond to FACS (blue), movie with all values (green) and movie with start values only (red, see below). All intensity ranges are normalized to $[0; 1]$ as FACS and movie analysis have different units. Although the FACS and movie lines do not exhibit striking similarity, the general shape is of importance. Common

features are a prominent peak in the right half of the distribution and the trailing-off to the left without a second peak. The skewness of the movie distribution is -1.51, the kurtosis is 6.71, the mean is 0.07 and the standard deviation is 0.53. Figure 3.7(b) also shows the Nanog distribution, but here only the first three measurements of each cell are considered to eliminate cell cycle effects as the Nanog intensity increases with the cell size. The distribution attributes are similar, with skewness -1.60, kurtosis 7.27, mean -0.09 and standard deviation 0.52. This effect cannot be eliminated in the FACS data as the position within the cell cycle of the cells is not known.
An interesting feature is the close similarity of the distributions in 3.7. This is suggested visually (see also the green and red line in Figure 3.6) and their relative entropy (or Kullback-Leibler divergence) is 0.021, indicating close relationship. The shift of the mean to the left from (a) to (b) can be explained by the increase of Nanog during the cell cycle, which is not shown in (b). But the variance is almost equal, showing that the start values alone are capable of producing an intensity spectrum of the same width. This proves that the cell cycle-dependent increase of Nanog abundance is not responsible for the general shape of the distribution, but that the skewness or heterogeneity is rather an intrinsically regulated feature of mESCs. In [27] and subsequent modeling approaches ([60, 78], see below) this fact is completely neglected.
The Nanog expression time courses of all 31 trees, from which the histograms are generated, can be seen in Figure 3.8. The plots show the starting intensity values of each generation.
The Nanog intensities extracted from the movies have arbitrary units, between 0 (no Nanog) and 14.66 (very high Nanog) on a linear scale, resulting from the quantification process which simply sums up the intensities of the cell pixels in the nucleus and subtracts the background. However, from the similar shape of the FACS and the image Nanog distributions, a relation between image intensities and true Nanog counts could be established by a mean comparison: there are on average 1E6 Nanog monomers in ESCs (see above), and the mean intensity value of all quantified cell timepoints is 1.1792 (linear scale) in the movies, resulting in a factor of 848,030 Nanog molecules per image intensity unit. However, the mean relation factor is a very coarse measure and the measurement errors in each method are not regarded.

3.3.3 Reestablishment of the steady state Nanog levels

The FACS cell sorting experiment described above (3.3.1) yields insights on the population level only. However, it would be interesting to know how the reestablishment of the original Nanog intensity distribution (if there is one) takes place at the single cell level. There are several questions that can be asked: (i) how long does the process take?, (ii) can we see the same speed bias as in the FACS data, meaning that LN cells take longer than HN cells to restore it?, (iii) how volatile are single cells in Nanog expression between generations?, and (iv) can the restoring process be explained by this general volatility, or is it rather the result of some destined cells that bear a long-term program for change?

As the cells in the movies are a random sample of the complete distribution, the sorting needs to be done in a post-hoc fashion. To execute this cell sorting cells from generation 1 (the daughter cells of the first division event in a tree) are used to group the trees according to their intensity values. Generation 0 (the first cell) is not utilized here because its state within the cell cycle is not known. Unfortunately, there are no trees whose generation 1 cells fall into either the dimmest two percent or the brightest five percent range, the gates used

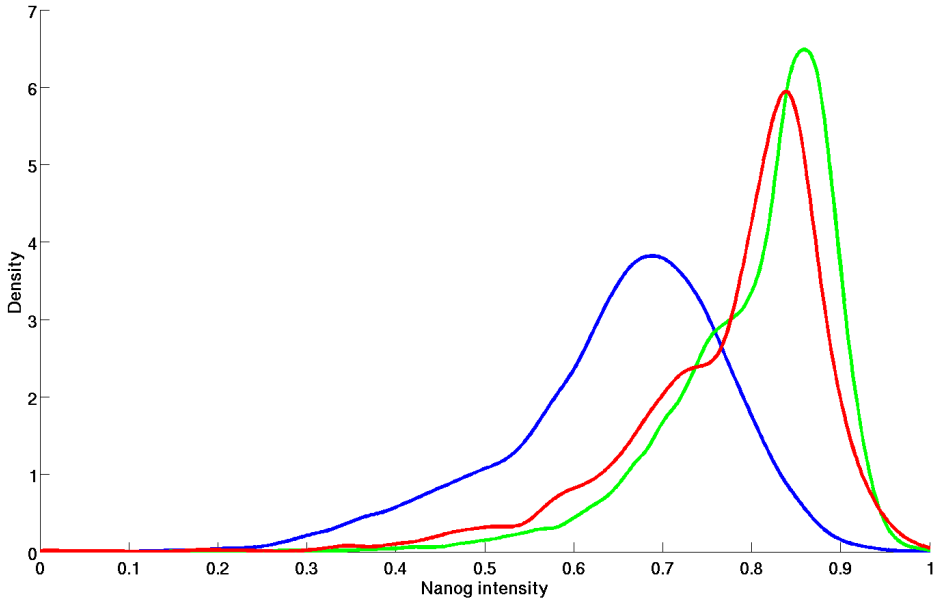
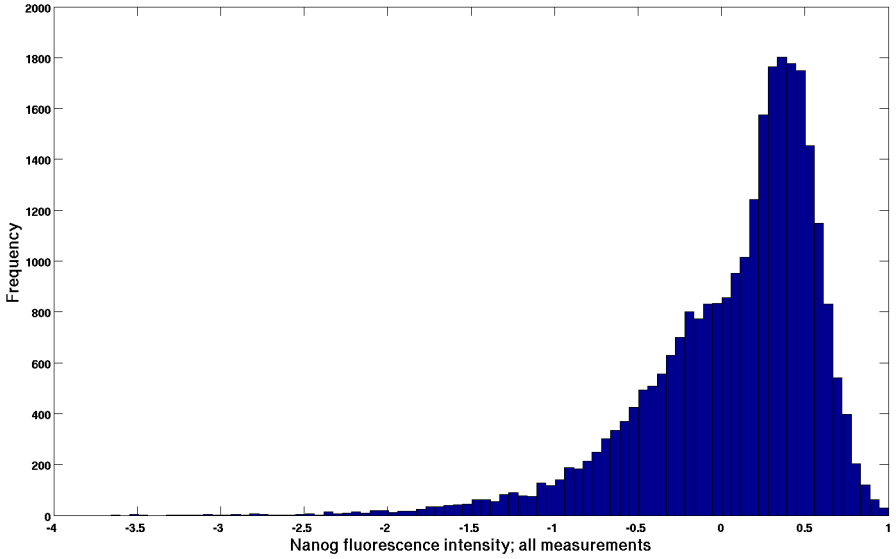
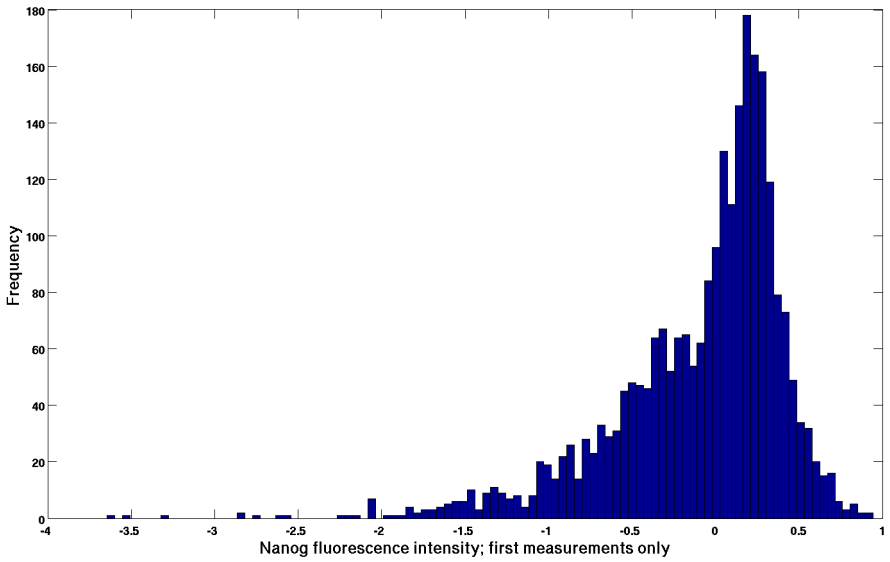


Figure 3.6: Nanog intensity distributions of the FACS population data (blue line), the time-lapse movies with all measurements (green line) and the time-lapse movies with start values only (red line). The x-axis shows the Nanog intensity (normalized to $[0; 1]$) and the y-axis the frequency. The curves are kernel density estimations of the underlying histograms.

for the FACS sorting experiment. Thus the sorting of the movie cells is performed slightly differently, which eventually might cause differences to the FACS results. The observation of FACS sorted cells (2% dimmest and 5% brightest) in time-lapse movies is currently underway in the ISF laboratory. The Nanog intensity distribution of the start measurements (to exclude cell cycle effects) is split into percentiles or *compartments*. A cell belongs to a certain compartment if its measured intensity value is within the boundaries of this compartment. Figure 3.9(a) shows the compartments of the distribution and the number of cells in each of them. We consider cells to lack Nanog completely if they are within the lowest ten percent, and to have a large abundance of Nanog if they are within the highest ten percent of the intensity distribution. The boundaries of the compartments which are used to group the trees according to their generation 1 Nanog intensities are thus $0.1, 0.25, 0.5, 0.75$ and 0.9, where the 25-, 50 and 75-percentiles are introduced to split up the center of the distribution for studying the cell dynamics in there in a more detailed way. The number of trees in each compartment is given in Figure 3.9(a). Table 3.3 shows the repopulation analysis in detail, providing the number of cells in each compartment over four generations. A visualization of this table as pie diagrams is shown in 3.10, and an interpolation of the minimum number of generations until the full reestablishment of the distribution is pictured in Figure 3.11. The first column of the table is the compartment number and the second column shows its boundaries. Column three lists the number of trees in each compartment. Column four shows the fraction of cells in generations $g \in \{1, 2, 3, 4\}$, with one generation per row. Each entry in the probability vectors represents one compartment. The fifth column shows the difference



(a)



(b)

Figure 3.7: Nanog intensity distributions. **(a)** The Nanog intensity distribution of all cells in generations 0 to 4. Values on the x-axis are log-scale and arbitrary but have a linear correlation with the FACS intensities (see text). The y-axis denotes frequency of the events. All measurements of the cells are included. **(b)** The Nanog intensity distribution of all cells in generations 0 to 4. Here, only the first three measurements of each cell are included to eliminate cell cycle-effects. Axes are equal to (a).

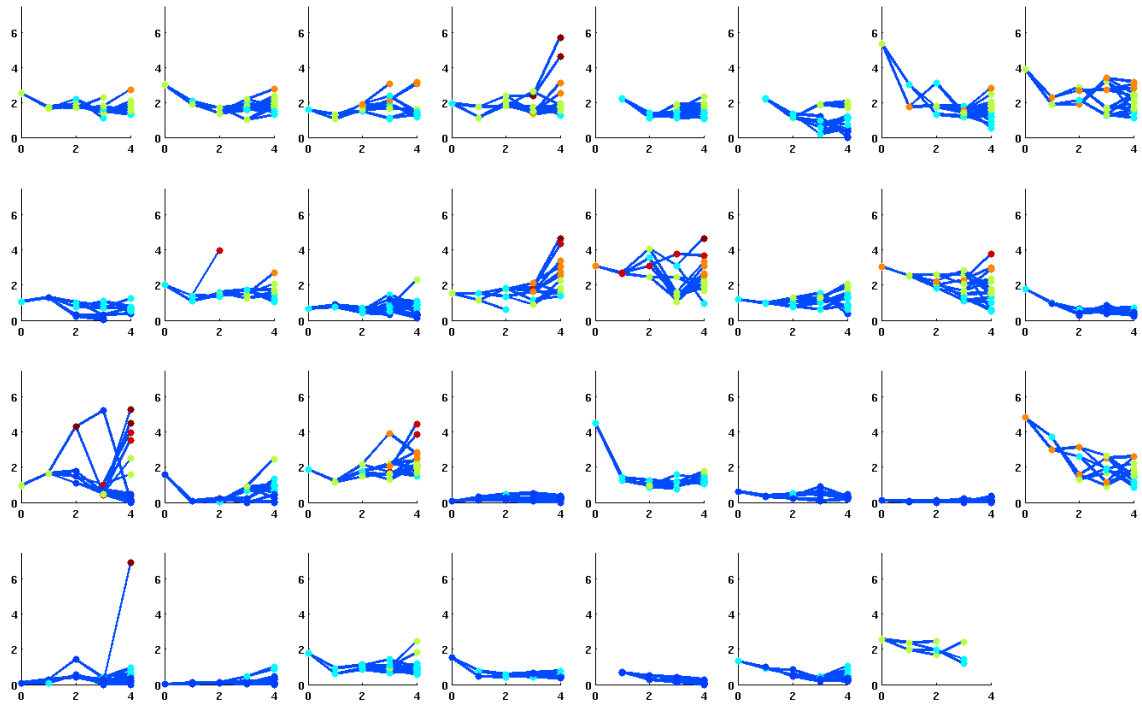
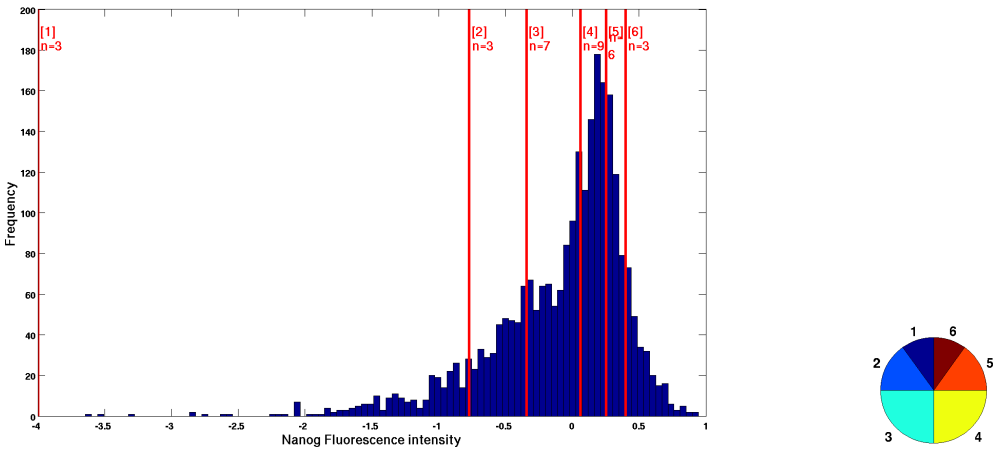


Figure 3.8: Time courses for all trees. The y-axis shows the Nanog expression level of the start values (linear scale), the x-axis the experiment time in generations. The color of the points corresponds with the expression level (from blue = low to red = high).

d_g of the frequency vector f_g in generation g to the original distribution r provided in the first row, calculated as the sum of differences $d_g = \sum_{i=1}^{n} |f_g(i) - r(i)|$, with $n = 6$ the number of compartments. The last column shows the Kullback-Leibler divergence KL_g of the vectors f_g and r, which is an established measure of the distance between two distributions.
Generation 1 cells by definition fall completely into their sorted compartment, and with each further generation the distribution of cells in the different compartments approaches the original Nanog distribution. Both distance measures d_g and KL_g exhibit the same trends, decaying rapidly in the extreme compartments, but staying rather constant after an initial decrease in the middle ones.
The best correspondence between the original distribution and a sorted one is observed in generation 4 of the third compartment, where $d_g = 0.4$ and $KL_g = 0.1$. Although the reestablishment is not yet completed here (probably because the movies last only for four generations), it supports the idea that cells with medium expression level have a stronger tendency to repopulate the complete Nanog distribution due to their flexibility in Nanog expression (see below and Figure 3.12). It is also clear that cells from the tails of the distribution need to traverse to the respective opposite end of the distribution, which necessarily takes more time than spreading out from the middle to both extremes simultaneously.
Summing up, question (i) about the duration of the reestablishment cannot be fully answered due to the censored data, but the data indicate that already after three rounds of division (from generation 1 to 4), a subset of the cells is capable of almost reproducing the original distribution.



(a) (b)

Figure 3.9: Definition of the intensity distribution compartments. **(a)** The histogram shows the distribution of Nanog intensities of the first cell measurements (x-axis is the intensity, y-axis the frequency). The red vertical lines represent the compartment boundaries and each compartment (noted in square brackets) spans between two red lines. For example, compartment 1 contains cells with intensities between -4 and -0.77, corresponding to the 0 and 0.1-percentiles. The number of trees n in each compartment is given. **(b)** The pie diagram shows the relative amount of each compartment, and is used as a reference in Figure 3.10.

Question (ii) about the rate bias in LN and HN cells equally struggles with the difficulty that the movies are too short to observe a complete reproduction. An interpolation of the measured differences KL_g (Figure 3.11) shows that the low and middle cells would need about five to six generations to restore the complete distribution (assumed if $KL_g < 0.01$), while the high and extreme high cells would need longer, up to 13 generations. The fastest restoring can be seen in compartment 3. This is valid only if the KL divergences decay exponentially, which seems to be a reasonable assumption for compartments 1, 5 and 6, but not so much for the others. Taken together, we can conclude that the high cells need longer to reestablish the original distribution, contradictory to the FACS results from above. The current experiment in the ISF laboratory (see above) might bring clarification on this issue by deciphering whether one of the former experiments ran awry or whether there is a real difference between the 2/5% sorted and the 10% sorted cells.
Question (iii) about the expression flexibility or volatility is best answered by the plot in Figure 3.12. It depicts the range of fluorescence intensities of the two daughter cells (y-axis), given that the current cell expresses Nanog at a certain level (x-axis). Low cells are observed to give rise to low cells preferentially and high cells tend to divide into high cells, too, while cells with an intermediate Nanog expression exhibit a broad spectrum of intensities in the next generation. This underlines the idea that cells from the middle of the distribution can repopulate the entire distribution faster than cells from the extremes.
The question (iv) about the mechanism of the distribution reestablishment is also answered by the volatility plot (Figure 3.12), showing that cells anywhere in the distribution are capable of producing offspring with slightly or heavily different expression levels. These cells

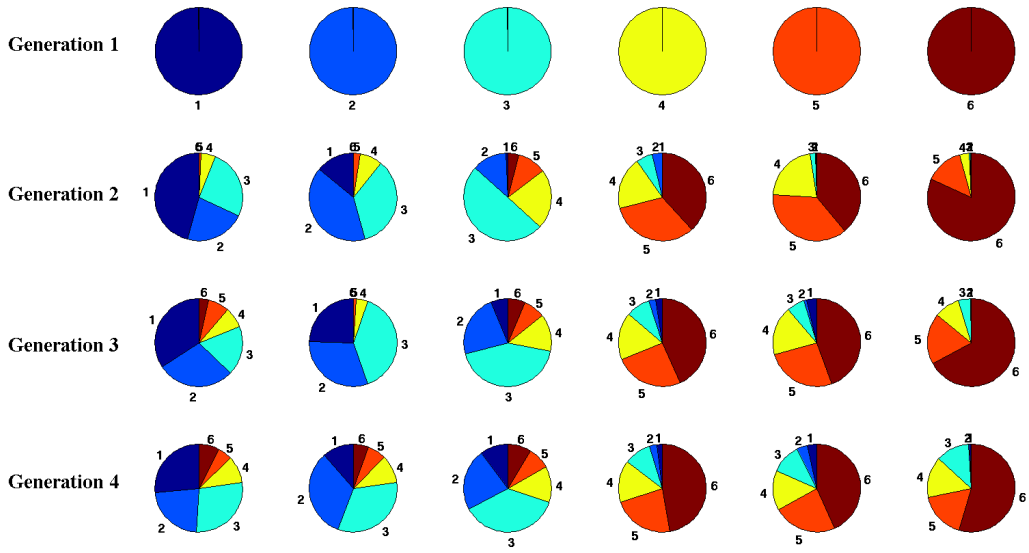


Figure 3.10: Reestablishment of the original Nanog intensity distribution. The pie diagrams visualize the numbers in Table 3.3 and show the fractions of cells within each compartment over generations 1 to 4. Each row represents one generation, each column one compartment. Moving from generation 1 to 4, the reestablishment can be observed by the changing fractions of cells in each compartment. The reference pie of the original Nanog distribution is given in Figure 3.9(b).

can then further split into daughters with distinct levels, until the steady state distribution is reached. Since this expression flexibility is not a rare property of few cells only (boxes are not very narrow), it can be stated that all cells bear the capability of restoring the steady state distribution of Nanog expression. Nonetheless, the timing of this process can substantially diverge, depending on the type of cells selected (extremely changing or not). The time scale of restoring is addressed in Figure 3.11 (see above) and the analysis of the expression memory below.
Another interesting aspect is whether the cells communicate to induce this diversification in Nanog intensities. This is not addressed here, but could be examined by observing whether cells in clumps behave differently than isolated cells, or by measuring signal molecules in the medium surrounding the cells. Current experiments in the laboratory of Ian Chambers suggest an effect of the plating density on the Nanog intensity (personal communication, unpublished data). If the assumption of LN cells taking longer to restore the skewed distribution (from the FACS experiment) turns out right, this might partly be due to the fact that LN cells are a mixture of ESCs and cells already committed to differentiation - which cannot give rise to HN cells anymore.

In the following we assume that all cells randomly picked from a population have the potential to restore the complete spectrum of Nanog expression. This is sound in the light of the six weeks FACS sorting experiment (data not shown) and the (interpolated) single cell data.

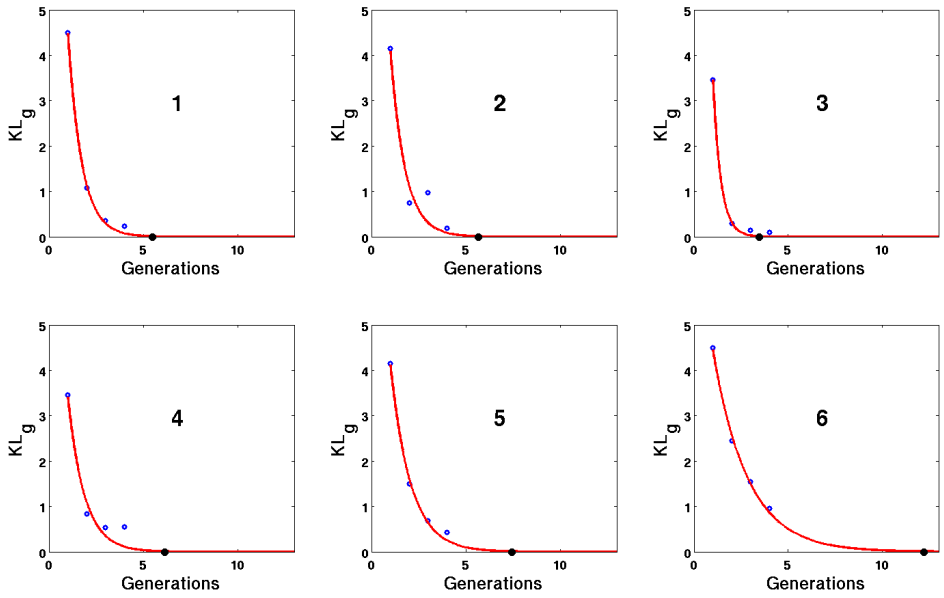


Figure 3.11: Generations before the full reestablishment of the original Nanog distribution. Each subplot represents one compartment indicated in the plots. The x-axis denotes generations, the y-axis the KL_g distance measure. The blue points are the measured values (see Table 3.3), the red line is a fitted exponential function. The black ball on the x-axis denotes the minimum number of generations where $y < 0.01$, thus indicating the estimated waiting time until the original Nanog distribution is restored.

3.3.4 Division and cell cycle characteristics

In Figure 3.13(a), the fluorescence loss l during cell division is shown. The loss of intensity for a cell c with daughter cells $f1$ and $f2$ is calculated as $l = \frac{i_l(c)-i_s(f1)-i_s(f2)}{i_l(c)}$, where $i_l(x)$ denotes the average of the last three measured intensities of cell x, and $i_f(x)$ similarly the average of the first three measured intensities (averaging was done to account for singular measurement errors). In an ideal division, the intensities of the daughter cells sum up to the intensity of the mother cell, resulting in a loss $l = 0$. However, as shown in the plot, the loss distribution is shifted to the right with an average loss of 8%, suggesting a rather messy cell division. Current experiments in the ISF laboratory indicate that Nanog is transmitted from the nucleus to the cytoplasm, which can account for the loss in some divisions since the quantification in the movies takes place in the nucleus only. Whether this Nanog transport is active, passive, or completely unregulated remains an open question. A gain in fluorescence (negative values of l) is unexpected, but is possibly due to small measurement errors that result in large relative losses in very dim cells (near zero intensity), owed to the calculation of the loss. Another possible explanation is motivated by the larger prevalence of gains in high cells: it could be possible that the Nanog production rate in these cells is high enough to add proteins to the amount received during division until the next observed timepoint. In Figure 3.13(b) the relative fluorescence loss is shown with respect to the mother cell intensity.

Comp.	Bounds	Trees	Generation 1/2/3/4	d_g	KL_g
ALL		31	$r = [0.10; 0.15; 0.25; 0.25; 0.15; 0.10]$	$d_0 = 0$	$KL_0 = 0$
1	$0.0 - 0.1$	3	$f_1 = [1.00; 0.00; 0.00; 0.00; 0.00; 0.00]$	$d_1 = 1.80$	$KL_1 = 4.49$
			$f_2 = [0.46; 0.22; 0.26; 0.05; 0.01; 0.00]$	$d_2 = 0.88$	$KL_2 = 1.08$
			$f_3 = [0.34; 0.29; 0.18; 0.07; 0.08; 0.04]$	$d_3 = 0.76$	$KL_3 = 0.36$
			$f_4 = [0.26; 0.23; 0.28; 0.10; 0.05; 0.08]$	$d_4 = 0.54$	$KL_4 = 0.23$
2	$0.1 - 0.25$	3	$f_1 = [0.00; 1.00; 0.00; 0.00; 0.00; 0.00]$	$d_1 = 1.70$	$KL_1 = 4.15$
			$f_2 = [0.14; 0.40; 0.35; 0.08; 0.02; 0.00]$	$d_2 = 0.78$	$KL_2 = 0.74$
			$f_3 = [0.25; 0.31; 0.39; 0.04; 0.01; 0.00]$	$d_3 = 0.89$	$KL_3 = 0.97$
			$f_4 = [0.11; 0.33; 0.33; 0.10; 0.07; 0.06]$	$d_4 = 0.54$	$KL_4 = 0.19$
3	$0.25 - 0.5$	7	$f_1 = [0.00; 0.00; 1.00; 0.00; 0.00; 0.00]$	$d_1 = 1.50$	$KL_1 = 3.46$
			$f_2 = [0.01; 0.13; 0.50; 0.22; 0.11; 0.04]$	$d_2 = 0.50$	$KL_2 = 0.30$
			$f_3 = [0.06; 0.23; 0.43; 0.14; 0.08; 0.07]$	$d_3 = 0.51$	$KL_3 = 0.14$
			$f_4 = [0.10; 0.23; 0.37; 0.13; 0.08; 0.09]$	$d_4 = 0.40$	$KL_4 = 0.10$
4	$0.5 - 0.75$	9	$f_1 = [0.00; 0.00; 0.00; 1.00; 0.00; 0.00]$	$d_1 = 1.50$	$KL_1 = 3.46$
			$f_2 = [0.00; 0.04; 0.06; 0.20; 0.32; 0.38]$	$d_2 = 0.92$	$KL_2 = 0.84$
			$f_3 = [0.02; 0.02; 0.09; 0.17; 0.25; 0.43]$	$d_3 = 0.88$	$KL_3 = 0.53$
			$f_4 = [0.02; 0.03; 0.10; 0.16; 0.22; 0.47]$	$d_4 = 0.89$	$KL_4 = 0.54$
5	$0.75 - 0.9$	6	$f_1 = [0.00; 0.00; 0.00; 0.00; 1.00; 0.00]$	$d_1 = 1.70$	$KL_1 = 4.15$
			$f_2 = [0.00; 0.00; 0.02; 0.22; 0.37; 0.39]$	$d_2 = 1.02$	$KL_2 = 1.49$
			$f_3 = [0.04; 0.01; 0.07; 0.18; 0.26; 0.44]$	$d_3 = 0.91$	$KL_3 = 0.69$
			$f_4 = [0.04; 0.04; 0.11; 0.14; 0.23; 0.43]$	$d_4 = 0.84$	$KL_4 = 0.43$
6	$0.9 - 1$	3	$f_1 = [0.00; 0.00; 0.00; 0.00; 0.00; 1.00]$	$d_1 = 1.80$	$KL_1 = 4.49$
			$f_2 = [0.00; 0.00; 0.01; 0.04; 0.14; 0.82]$	$d_2 = 1.44$	$KL_2 = 2.45$
			$f_3 = [0.00; 0.00; 0.04; 0.09; 0.19; 0.67]$	$d_3 = 1.22$	$KL_3 = 1.54$
			$f_4 = [0.01; 0.00; 0.12; 0.15; 0.17; 0.54]$	$d_4 = 0.93$	$KL_4 = 0.96$

Table 3.3: Reestablishment of the original Nanog intensity distribution in the single cell data. The trees in the movies are sorted post-hoc into six compartments (column 1) by the starting intensities of its generation 1 cells. The compartment bounds are listed in the second column. The third column gives the number of trees in each compartment. Column four shows the relative frequency of cells in generations 1, 2, 3 and 4 (one per row) respectively as frequency vectors, each entry representing one compartment. Column five shows the difference d_g of the frequency vector f_g in generation g to the original distribution r provided in the first row, calculated as $d_g = \sum_{i=1}^{n} |f_g(i) - r(i)|$. The last column shows the Kullback-Leibler divergence KL_g of the vectors f_g and r. Both distance measures exhibit the same trends, decaying rapidly in the extreme compartments, but staying rather constant after an initial decrease in the middle ones. A visualization of this table in various ways can be found in Figures 3.10 and 3.11.

The plot indicates that higher losses are more prevalent in low cells, although high cells also exhibit a substantial range of loss values.

Figure 3.13(c) shows the asymmetry of Nanog allocation between the nascent daughter cells in a division. In an ideal division both daughter cells receive equal amounts of Nanog, but as cells neither are completely homogeneous nor show perfect divisions (see loss above), a biased repartition of a protein is actually something to be expected. Here we measure the asymmetry $a = |\frac{i_f(f1) - i_f(f2))}{i_f(f1) + i_f(f2)}|$, and a perfectly symmetric division would result in $a = 0$.

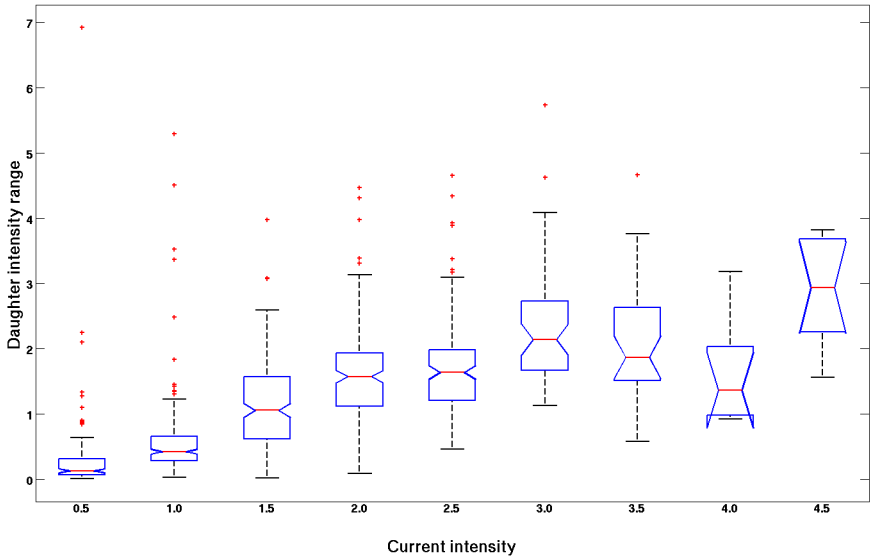


Figure 3.12: Nanog expression volatility. The x-axis shows the intensity of the current cell (binned in 0.5 ranges), the y-axis the intensity range of the daughter cells. The notches in the boxes indicate the 95% comparison interval of the medians, showing that two medians are significantly different if the notches do not overlap. A 1-way ANOVA test to compare the means yields a p-value of 0, supporting the idea of differences between the daughter cell intensity ranges. The strange notch at intensity 4.0 is a consequence of few data points in this section as the notch is larger than the box itself.

The absolute value is taken since the assignment of $f1$ and $f2$ is arbitrary. The mode of the distribution is zero and it follows the right half of a shifted and scaled t-distribution, indicating frequent, but rather slight asymmetric divisions. The idea of asymmetry being responsible for heterogeneity has already been explored in several studies, indicating the adequacy and importance of quantifying asymmetry (see [73] for the difficulty of separating segregation effects from transcriptional noise, and [71] for a comprehensive review). In Figure 3.13(d), the division asymmetry is plotted against the Nanog intensity. The cells exhibiting the most asymmetric divisions are the low ones, but the asymmetry is a feature of cells with all intensities. Similar to the fluorescence loss, the higher asymmetry in low cells could be due to technical noise (measurement errors) and the definition of a.

We propose that the existence of fluorescence loss during division and the asymmetric segregation of Nanog proteins are inherent features of the utilized ESCs. An accurate observation and quantification in a movie with short time intervals between pictures could help to clarify the role and biological significance of these two processes. Current detailed experiments in the ISF laboratory confirm the existence of loss and asymmetry in divisions, but need further analysis.

An interesting question now is whether the loss of Nanog during division or the asymmetric allotment of Nanog is responsible for the skewness of the intensity distribution. This is analyzed with a new molecular model (see modeling section 3.4.3) and yields that the asymmetry is an essential mechanistic feature of the heterogeneous expression of Nanog, while the loss is not as important but should not be neglected either.

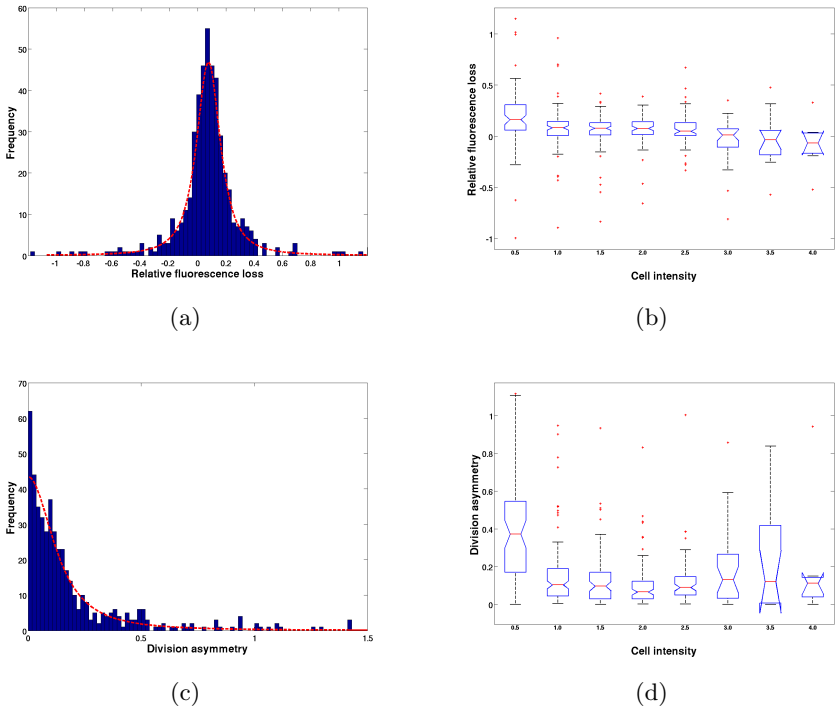


Figure 3.13: Cell division characteristics. **(a)** Loss of fluorescence during cell division. The x-axis denotes the fluorescence intensity loss in relation to the mother cell intensity (see text for calculation), the y-axis shows the absolute frequency. Negative loss values indicate actual fluorescence gain. The red line is a shifted and scaled t-distribution with $df = 1.65$, $\mu = 0.08$, $\sigma = 0.09$, and a KS test yields $p = 0.97$. **(b)** Relative fluorescence loss, split by Nanog intensity. The x-axis denotes the expression level, the y-axis the range of relative fluorescence loss over all cells. **(c)** Asymmetry of Nanog allotment during division. The x-axis denotes the asymmetry (see text for calculation), the y-axis shows the absolute frequency. The red line is the right half of a shifted and scaled t-distribution with $df = 1.47$, $\mu = 0.01$, $\sigma = 0.12$, and a KS test yields $p = 0.1$. **(d)** Similar to (b), the division asymmetry split by Nanog intensity. The strange notch at 3.5 is due to few data points and extends beyond the box' width.

The average cell cycle time (from birth until division) over all cells is 15.85 hours. However, one can ask if cells with low Nanog expression proliferate with the same rate as those with high Nanog levels. The cell cycle times are measured in each compartment separately and are shown in Figure 3.14 as violin plots. It indicates that the cells proliferate equally fast, independent of the Nanog intensity. A Kruskal-Wallis test (data are not normally distributed) to compare the medians yields a p-value of 0.28, supporting this assumption. A Levene's test to compare the variances yields a p-value of 0.018, indicating different variances across the compartments, an idea also supported by the different widths of the distributions.
It is tempting to interpret the median equality as natural, but it somehow contradicts present ideas of defining ESCs. Usually, these exhibit rapid cell cycle times of approximately 12 hours [127], which increase with the level of differentiation. Thus as Nanog is considered an ESC specific protein, the low cells (which are more prone to differentiation [27]) could be expected to show slower proliferation rates than the high cells. This is not the case in

our data, indicating an undifferentiated ESC nature even in the low cells - as Chambers and co-workers already suggested [27]. Following this point of view, the difference between low and high cells is not in their pluripotent capacity, but in the decreased stability of the low state since these cells provide 'windows of opportunity' [27] for differentiation. This is in accordance with the large range of life times in the lowest compartment (tails are heavier), indicating the existence of a substantial amount of slowly proliferating cells - which possibly have 'the window open'. Note that outliers in the high compartments can be indicative for quiescent cells (transiently entering into G_0).

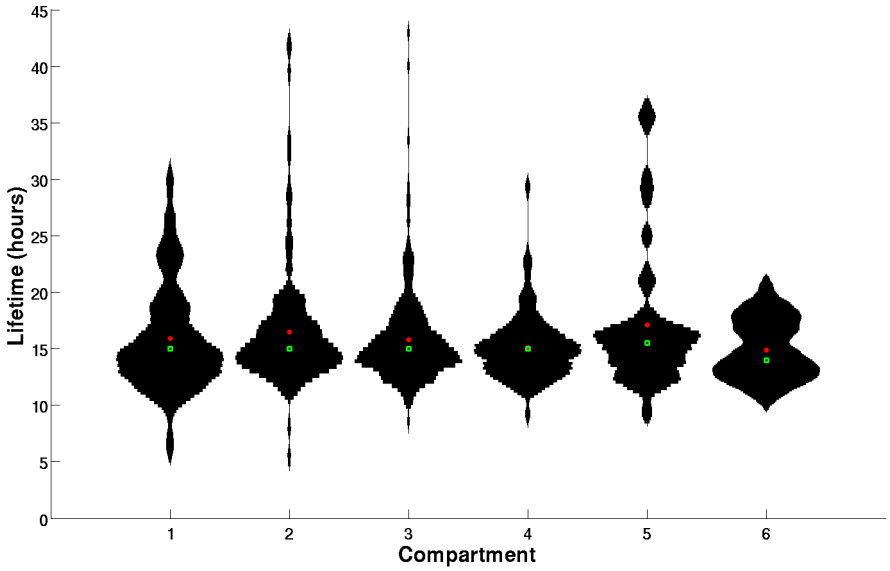


Figure 3.14: The cell cycle times of cells with different fluorescence intensities as violin plots. The x-axis denotes the compartments of the cells, the y-axis shows the life time in hours. Outliers are more abundant in the middle range and possibly stem from cells that transiently enter into G_0 during their cell cycle. The distribution itself is broader in the lowest compartment, however. The green squares denote the median of each distribution, the red circles the mean. The bimodality in the sixth compartment is possibly due to few data points. For details about the intensity compartments, please refer to the text and to Table 3.3.

3.3.5 Influence of the cell history

One can also ask if the cells have a memory, meaning that their Nanog intensity is not independent from their history. To treat this question formally one can ask whether the cells fulfill the Markov property. This property states that the chance of producing a certain level of Nanog expression in the daughter cells depends on the current expression state only, and not on the previous ones [11].
The cells are split into the same six compartments as defined above (see Table 3.3) and the rates of change between these compartments are measured by counting the transitions over one generation. Again, only the first measurements of each cell are considered to avoid any cell cycle effects. The change rates between all compartments without regarding the history

are listed in Table 3.4, which is colored as a heatmap. The differences between the change vectors for each compartment are calculated as average KL divergence and provided in the last column ($KL_i = \frac{1}{n-1} \sum_{j=1}^{n} KL(i,j)$ with i the current compartment, j all others and n the number of compartments). The data reflect the findings from above that low cells tend to produce low cells and high cells tend to produce high cells, while cells with medium Nanog expression generate daughter cells with all levels of Nanog expression. This explains why the middle cells are found to be fastest in restoring the complete Nanog distribution.
In order to check whether the ultimate cell history (the intensity of the mother cell) has an influence on these rates three-dimensional change rates (mother cell, current cell, daughters) are calculated. The result is shown in Table 3.5 in heatmap colors. If the change rates do not depend on the cell's history, the rates of one compartment to another should be equal for all histories (the rows in one group should be similar) and similar to Table 3.4. This is not the case in our data. However, the rates tend to be more similar in currently low or high cells than in currently middle cells, indicated by the lower average KL divergences in the extremes (calculated as above, provided in the last row). Consider, for example, the cells which are now in compartment 3 but have different histories: the rates change dramatically, depending on the ultimate cell development. Interestingly, in this compartment the diagonal tends to contain the highest values, suggesting that the cells prefer to go back to where they come from.
Note that there is a slight discrepancy between this table and Table 3.4 since in the latter one there are, for example, cells which change from 6 to 1, which are not present here. This is due to the limited number of generations in the trees which contain cells that have such a history but no quantified daughter cells. Therefore they do not occur here (as this table spans over three generations).
Concluding, this suggests that the cells know about their origin in general and that changes in Nanog expression are usually not a quick-response program that acts within one generation but at least partly involves a longer lasting decision. The diagonal in the third compartment and the preference of more extreme cells to stay extreme support this conclusion. A possible mechanism generating this pattern could be an epigenetic programming for change, like chromatin rearrangements or histone modifications (see [152] for an example of such a fluctuation-generating mechanism). Experimental strategies to validate this supposition include longer time-lapse movies (at least eight or nine generations) to quantify the loss of memory over time, or the time-resolved inspection of the Nanog promoter or histone methylation state in cells that change their Nanog expression level.

3.3.6 The existence of subgroups

In the previous analyses the distribution has been split into six distinct compartments of Nanog expression. The borders between the compartments have been motivated by the definition of the FACS gates to define cells with either very low or very high Nanog expression and by the idea to subdivide the mass of the cells that have a medium expression into four subgroups. Nonetheless, the number of 6 has been chosen arbitrarily, and the question is posed whether this compartmentalization is biologically meaningful or if it is more adequate to have a different number of compartments or none at all.
In our view a compartment of cells is biologically reasonable if it represents a true subgroup of cells with a distinct level of Nanog expression. Such a subgroup cannot be a rigid cage, evidenced by the volatility of Nanog expression over one generation (see Figure 3.12), allowing

Compartment	Number of cells (total = 767)	Daughter intensity (compartment) 1	2	3	4	5	6	KL_i
1	87	0.62	0.26	0.10	0.00	0.00	0.01	0.35
2	103	0.27	0.44	0.26	0.02	0.00	0.01	0.09
3	220	0.05	0.21	0.45	0.19	0.09	0.02	0.09
4	201	0.01	0.04	0.18	0.40	0.26	0.11	-0.10
5	96	0.00	0.00	0.10	0.42	0.28	0.20	0.10
6	60	0.03	0.00	0.08	0.28	0.22	0.38	-0.10

Table 3.4: Change rates from any compartment to any other without regarding the history of the cell. The numbers 1 to 6 in rows and columns denote the compartments as listed in Table 3.3. The rows represent different current intensity compartments, and columns three to eight represent different daughter cell compartments. The first column shows the current intensity compartment and the second one the number of cells in there. The last column provides the average KL divergence of the frequency vector to all others.

cells to wander between different protein abundances. But on the other hand it should be at least partly stable, defined by the probability that a cell exhibiting the distinctive level of Nanog expression of this group stays within it over some generations. In other words, a compartment x is defined as stable (and thus of biological interest as a putative subgroup) if the chance of cells in this compartment to stay there in the next generation as well (denoted by P_{xx}) is greater than a certain threshold. This threshold can be defined in various ways, yielding different levels of stability. The most relaxed way is to define a compartment as stable if $P_{xx} > 0$ holds. Taking a look at Table 3.4 (diagonal entries), all six compartments are stable with this definition. A second way is to require P_{xx} to be greater than all other change rates, meaning a cell rather tends to stay in its own compartment than to roam into another compartment y, thus $P_{xx} \geq P_{xy}, y \in \{1 \dots n\} \setminus \{x\}$. With this definition all compartments are stable except compartment 5, where cells preferably change to compartment 4. A further possibility, and the most rigid one, is to define stability by $P_{xx} \geq 0.5$, meaning the cells have the highest chance to stay in their compartment rather than changing it. By using this form only compartment 1 (very low expression) is stable.
Neglecting the relaxed stability definition ($P_{xx} > 0$) because of its indistinctive nature, the most meaningful number of compartments can be derived from the analysis of the two plots in Figure 3.15. In both the number of compartments is shown on the x-axis, and the number of stable compartments on the y-axis. The compartment bounds are now defined by equal quantile sections of the complete Nanog distribution. For example, the bounds in the case of four compartments correspond to the quantile values of $[0; 0.25; 0.5; 0.75; 1]$. In 3.15(a), the stability is defined rigidly by $P_{xx} \geq 0.5$, while in 3.15(b) a compartment x is stable if $P_{xx} \geq P_{xy}, y \in \{1 \dots n\} \setminus \{x\}$. The maximum number of stable compartments with the rigid definition is two and it goes down to zero for 20 or more compartments, while in the less conservative case the number slowly increases with the compartment count. The red points show the fraction of stable compartments, which has its peak at two compartments and decays for both definitions until it is virtually zero after 14 or 17 compartments, respectively. From the left plot it could be concluded that there are only two stable compartments, while the right plot suggests the existence of more subgroups. Considering the very conservative nature of the first (and the decreasing chance of scoring a probability > 0.5 in the case of many neigh-

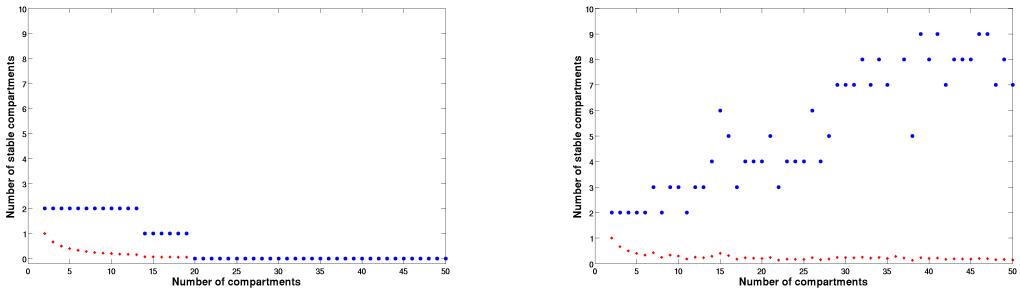
Mother	Current	Number	Daughter cells 1	2	3	4	5	6	KL_i
1	1	61	0.66	0.23	0.11	0.00	0.00	0.00	0.04
2		16	0.50	0.38	0.12	0.00	0.00	0.00	
3		6	0.67	0.17	0.00	0.00	0.00	0.17	
4		4	0.50	0.50	0.00	0.00	0.00	0.00	
5		0	-	-	-	-	-	-	-
6		0	-	-	-	-	-	-	-
1	2	14	0.43	0.43	0.07	0.07	0.00	0.00	0.04
2		25	0.48	0.24	0.24	0.04	0.00	0.00	
3		50	0.12	0.60	0.28	0.00	0.00	0.00	
4		14	0.29	0.21	0.43	0.00	0.00	0.07	
5		0	-	-	-	-	-	-	-
6		0	-	-	-	-	-	-	-
1	3	14	0.21	0.14	0.57	0.00	0.07	0.00	0.10
2		21	0.00	0.38	0.57	0.00	0.05	0.00	
3		105	0.03	0.28	0.48	0.19	0.02	0.01	
4		54	0.00	0.13	0.46	0.26	0.13	0.02	
5		20	0.10	0.00	0.15	0.35	0.35	0.05	
6		6	0.33	0.00	0.00	0.17	0.17	0.33	
1	4	0	-	-	-	-	-	-	0.11
2		2	0.00	1.00	0.00	0.00	0.00	0.00	
3		33	0.00	0.09	0.24	0.39	0.24	0.03	
4		70	0.03	0.03	0.20	0.40	0.26	0.09	
5		64	0.00	0.02	0.12	0.47	0.23	0.16	
6		32	0.00	0.00	0.19	0.31	0.34	0.16	
1	5	0	-	-	-	-	-	-	0.05
2		0	-	-	-	-	-	-	-
3		16	0.00	0.00	0.00	0.38	0.56	0.06	
4		32	0.00	0.00	0.16	0.44	0.22	0.19	
5		22	0.00	0.00	0.14	0.41	0.05	0.41	
6		26	0.00	0.00	0.08	0.42	0.38	0.12	
1	6	0	-	-	-	-	-	-	0.02
2		0	-	-	-	-	-	-	-
3		0	-	-	-	-	-	-	-
4		8	0.00	0.00	0.12	0.38	0.25	0.25	
5		12	0.00	0.00	0.08	0.25	0.25	0.42	
6		40	0.05	0.00	0.07	0.28	0.20	0.40	

Table 3.5: Change rates from any compartment to any other, regarding the ultimate history of the cell. The numbers 1 to 6 in rows and columns denote the distribution compartments as listed in Table 3.3. The rows represent different histories (mother cell) and current intensity compartments, while the columns represent different daughter cell compartments.

bors), Table 3.4 and the fact that any positive chance of staying in the same compartment is interesting by itself, we vote for the second definition and state that there is good reason to

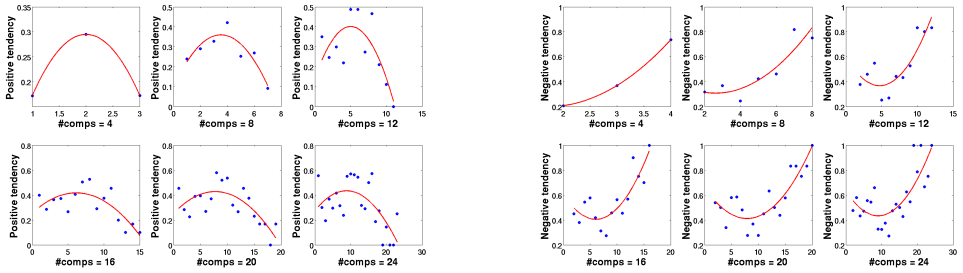
assume the existence of subgroups, albeit the exact number is hard to define.
In Figure 3.16 the change rates to higher (left column) and lower (right column) compartments are shown for a varying number of equally spaced compartments. Note that the six compartments above are not equally spaced and do not occur here. Nonetheless, the plots can be considered a continuous extension of Table 3.4. The positive/negative tendency is the sum of change frequencies to all higher/lower compartments, respectively. The individual plots in (a,b) are summarized in (c,d), where the compartment index is in relative terms of the total number. The bar plot in 3.16(e) merges (c,d) by plotting the positive tendency in blue (from (c)), the negative tendency in red (from (d)) and the remaining fraction in green (which is the chance to stay in the same compartment). The red lines in (a-d) are polynomials with degree two, having their minimum/maximum around 0.4 (relative compartment index) each. The equal location of the two extremes conveys that the middle cells tend to give rise rather to higher cells than lower ones, as previously indicated in Table 3.4 and Figure 3.12. This is conform with the generally larger abundance of high cells, obvious in the intensity histograms (Figure 3.7). The substantial negative tendency in the very low cells reflects the higher propensity of low cells to stay low, while the high negative tendency in high cells indicates an inclination to lower the expression. The fact that there is a peak or dip at all supports the idea that extreme cells are clearly different from middle cells and thus underlines the existence of distinct subgroups of mESCs, defined by their level of Nanog expression.
Taken together, the definition of stability and compartment boundaries is decisive for defining the number of stable compartments or subgroups. Thus we do not provide a definite number of subgroups but rather find that the assumption of their existence is reasonable in the light of the data. If, nonetheless, a fixed number is given, bootstrapping (omitting random cells from the analysis) would be an interesting approach to test the robustness of this number.

Note that, although the compartmentalization of the Nanog distribution is a good hint of the existence of subgroups, it does not explain how this distribution or division into different cell populations is actually generated. This is treated in the next section, where both previously published mechanistic models are evaluated in the light of the data and a new molecular model is proposed.



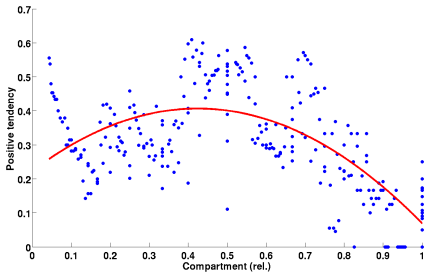
(a) (b)

Figure 3.15: Number of stable compartments vs. the number of compartments. Blue points denote the number, while red points denote the fraction of stable compartments in both plots. **(a)** Stability defined by the rigid property $P_{xx} \geq 0.5$. **(b)** Stability defined by the property $P_{xx} \geq P_{xy}$.

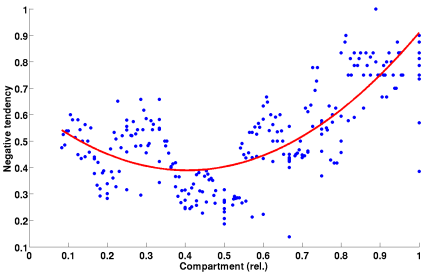


(a)

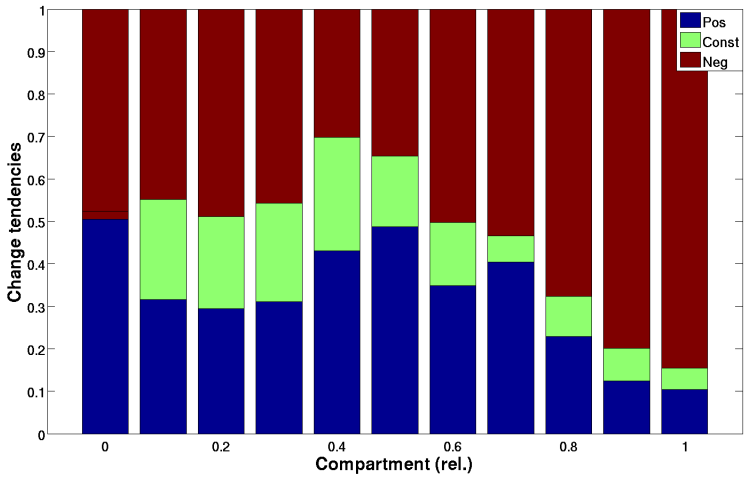
(b)



(c)



(d)



(e)

Figure 3.16: Compartment change rates for various numbers of equally spaced compartments. The red lines in (a)-(d) are polynomial fits with degree 2. The compartment indices in (c, d, e) are relative, with the highest number set to 1. **(a, c)** Sums of change rates to higher compartments (positive tendencies). Individual plots from (a) are summarized in (c), the maximum is at 0.41. **(b, d)** Sums of change rates to lower compartments (negative tendencies). Summary plot similar to the left column, the minimum in (d) is at 0.4. **(e)** Bar plot of (c, d) stacked. Blue bars denote positive tendencies, red bars negative ones and green bars the propensity to keep the expression level.

3.4 Molecular Nanog models

After the initial publication of the fluctuating Nanog levels by Chambers et al. [27] the question of how this fluctuation is generated at the molecular level has been posed. Few papers have tried to shed light on this problem so far, including one purely experimental approach [131], one purely theoretical modeling approach [60] and one study that combines both strategies [78]. The literature about fluctuations in gene expression in general has been fostered by the discovery of many genes with fluctuating expression levels in a macroscopically homogeneous cell state (see e.g. [27, 64, 66] for pluripotent cells). A comprehensive review of experimental and theoretical approaches to study non-genetic fluctuations in development is given in [71]. A series of papers by Raser and Paulsson deals with modeling the inherent stochasticity in gene expression, which are nicely reviewed in [134, 141]. In [178] the role of noise in generic developmental networks is assessed. The importance of understanding heterogeneity is outlined in a pharmacological context in [124, 143, 152].
In this section the theoretical models in [60, 78] are reviewed and their validity tested in the light of our data. Considering the results of this evaluation and the still unexplained features, a new and simple molecular model of Nanog expression is proposed which is able to explain parts of the previously disregarded features of ESCs.

3.4.1 The excitable system by Kalmar et al.

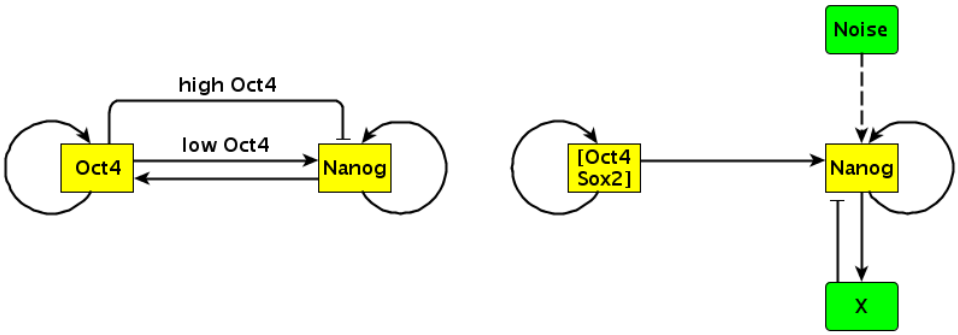
In [78] the authors propose a GRN (see Figure 3.17(a)) that consists of an excitable system with transcriptional noise. An excitable system is defined as a nonlinear dynamical system that can generate large responses to small stimuli if a certain threshold is passed. Usually the shape of the response is not dependent on the shape of the stimulus (i.e., nonlinear) [100]. A well-known example of an excitable system is the neuronal information propagation in the mammalian brain [102], where a sharp output peak (an action potential) is generated if the sum of the input signals is above a certain voltage threshold. In the Nanog model in [78] the cell is assumed to be in a steady state with high expression of Nanog (HN), and the input stimulus is generated by Gaussian white noise (mean 0 and normally distributed) in transcription. The response is a transient cell state with low Nanog expression (LN). Due to the nature of the excitable system, this state is not stable and the cell can either return to the HN state or lose its pluripotency and differentiate. The construction of their model is stimulated by two key observations: Nanog is fluctuating, as reported in [27], and high levels of Oct4 seem to repress Nanog in an unknown way [101, 125, 131]. The model is implemented as an ODE (ordinary differential equations) system, and chemical reactions are also simulated with a Gillespie algorithm.
The authors analyze the dynamic behavior of their cell model and postulate predictions whose validity they assess with laboratory experiments. Their conjectures state that (i) the LN state is unstable and returns to the HN state immediately, and (ii) the transition from HN to LN cells should be rare while the reverse transition from LN to HN should be frequent.
To prove these two suppositions they perform FACS sorting experiments similar to the ones described above (3.3.1) and time-lapse movies of their ESCs. The authors conclude from their data that the LN cells give rise to HN cells more readily than the other way round because the sorted populations exhibit different repopulation velocities.

However, their paper suffers from some inconsistencies in the data generation and interpretation. First, they do not correct for cell cycle effects - a cell that is classified as LN

could actually be a cell having just divided and which is on its way back to the HN state. Second, for the TNG4 (Nanog transcription reporter, see above) line no observations for the time beyond 48 hours are shown, and there could be a substantial change in the behavior after this time. Third, the image data they provide to show that LN cells indeed give rise to HN cells are not accurately quantified but only inspected by eye, which could be a source of errors. Additionally, the proof that the *majority* of LN cells behaves this way (as their model predicts) is not given. Fourth, there are no data supporting the second hypothesis - there is no reasonable background to assume that HN cells give rise to LN cells more rarely than vice versa if not all cells are surveyed and quantified. Last, the time course generated from their model shows the rapid and rare transitions between the LN and HN state they predict. Yet, averaging the residence times of cells in both states from this model actually does not generate the approximate 20/80 LN/HN distribution they refer to because the LN cells are significantly underrepresented.
An additional line of evidence speaks against their model conjured from our single cell data: the LN state is stable and cells in the LN state are not inevitably destined to revert to the HN state, clearly shown by the high number of cells that stay within the low compartment in Table 3.4. Furthermore, the change rates between the HN and LN states are not significantly different as they propose since the positive and negative change rates are rather equal and actually the low cells tend to stay low even more than the high cells tend to stay high. This is evidenced in Table 3.4 and Figure 3.16(e).
Summing up, the authors in [78] do not give a sound proof that the excitable model with Oct4 and Nanog is an adequate explanation for the fluctuations of Nanog, and our data rather suggest that it is either wrong or too simple. It should be noted that, indeed, there are excitable systems in biology, like the neurons (see above) or the transient differentiation into competence of *B. subtilis* [164], indicating that the idea itself is not futile at all. A similar mESC system (but not excitable) was studied in [33], where the repressive effect of high Oct4 levels on Nanog is executed by the intermediate player Gata6 (see also chapter 2).

3.4.2 The two models by Glauche et al.

In [60] the authors also propose two GRNs that could explain the fluctuating expression levels of Nanog in ESCs. The structure of the two distinct models is given in 3.17(b). Both models contain the Oct4-Sox2 heterodimer and Nanog. The first model is termed "fluctuation model" by the authors and additionally contains a noise component (with unspecified source) which influences the expression of Nanog. The second model ("oscillation model") does not incorporate noise, but introduces a third protein termed X which is connected with Nanog in a negative feedback loop: its expression is increased by Nanog, but X in turn downregulates Nanog. Having constructed these two distinct models, the authors analyze the behavior of each model separately and predict observations both on the population and the single cell level. With the help of these prognoses they suggest two experimental strategies that should produce different outcomes depending on the true model - thus an experimentator could easily define which model is correct, given the true measurements either in single cells or in a population. In their study the authors implement the system by ODEs with an additional noise term.
Interestingly, Glauche and co-workers initially neglect the well-known positive feedback of Nanog on the expression of Oct4 and Sox2 [101]. They argue that this omission is reasonable since Nanog levels fluctuate more extremely than Oct4 levels, which would not be possible



(a) (b)

Figure 3.17: Three (one in (a), two in (b)) models that try to explain the fluctuations in Nanog expression. Triangle arrows denote activation, while blunt arrows denote inhibition in both networks. **(a)** The GRN excitable system in [78]. Oct4 and Nanog are both auto-activating, Nanog has a positive feedback on Oct4 and low levels of Oct4 activate the expression of Nanog, while high levels repress it. **(b)** The two distinct GRN models in [60] united. Both models contain the Oct4-Sox2 heterodimer and Nanog. These proteins are auto-activating and the Oct4-Sox2 dimer has a positive effect on the expression of Nanog. In the first model ('fluctuation') Nanog is influenced by a 'Noise' term, which is not an explicit protein but rather an effect from different sources of stochastic fluctuations. The second model ('oscillation') contains the player X, which is a hypothetical protein that is activated by Nanog, and in turn represses Nanog.

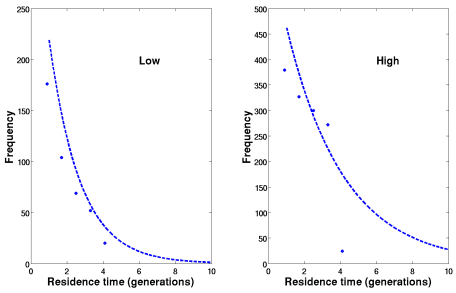
in a strictly positive feedback loop. Although this might be true for the inevitably limited scope of the model, it need certainly not be true for real cells. Additionally, there is an auto-inhibiting function of Oct4 reported in [131], which might serve exactly for the purpose of buffering the large fluctuations of Nanog. At the end of their study the authors re-insert the positive feedback of Nanog on Oct4 and Sox2 indirectly by a double inhibition, where Nanog prohibits the repressing effect of external differentiation signals on Oct4 and Sox2. This is, however, not exactly the same as the evidenced direct feedback as it potentially has different temporal lags and generates different expression patterns in the presence of further proteins. For the analysis of their models the Oct4-Sox2 dimer is not relevant once it has reached a steady state because it only serves as a source of constant activation of Nanog without feedback, and could thus be completely omitted.
From a dynamic analysis the authors in [60] conclude that the 20/80 distribution of LN/HN cells is reproducible by both models if appropriate parameters are chosen. These parameters are selected to match the results with previous findings about the skewed Nanog distribution. Thus, in the following it has to be considered that all their results are based on rather arbitrary parameter values.

The **oscillation** model contains an unspecified protein X (Figure 3.17(b)) which is solely responsible for the fluctuations in Nanog expression levels. Assuming that this model is true, and no noise in the Nanog gene expression pathway exists, a FACS sorting experiment (see also above) which separates LN and HN cells should produce characteristic repopulation trajectories until the original Nanog expression curve is reestablished. An "overshooting" is predicted, meaning that the sorted cells do not continuously shift towards the original expression distribution but rather swing as a whole entity with decreasing cohesion to the

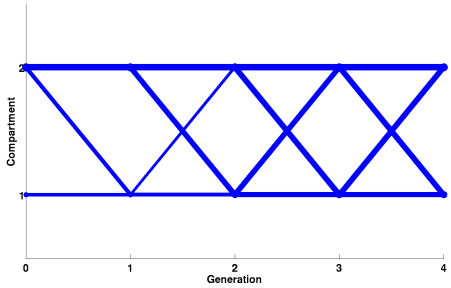
full reestablishment. For example, in the case of LN sorted cells, the cells would continue to change expression levels although the steady state is already reached. This would lead to transient states with higher numbers of cells in the opposing HN state, but the peaks of these fluctuations would decrease in the course of time until the steady state is completely stable. If there are single cell data present, a second experiment could be performed to assess the validity of the model. The residence times (time before a change of expression state) of cells in the LN state are predicted to be narrowly and symmetrically distributed around a low mean value, while the residence times in the HN state should be more spread out and have a slightly higher peak (but still rather symmetric).
Regarding the population data from above (Table 3.2 and Figure 3.5), we do not observe an overshooting during the reestablishment of the original distribution, rather a continuous decrease of the initially pure LN or HN cells, respectively, and an equally smooth increase of the opposite state cells. The residence times of the single cells seem to follow rather an exponential distribution than a symmetric one (Figure 3.18(a)). Additionally, the plots in Figures 3.18(b) and 3.18(c) show the transitions between different expression states and the frequency of staying constant (LN corresponds to compartments 1 and 2, while HN corresponds to 3-6). It can be seen that there are generally few transitions between the two compartments and oscillations (rapidly changing between compartments) almost do not occur. Therefore, although the analysis might be skewed as the single cells are observed only for four generations, the three results suggest that the oscillation model might be inadequate or too simple.

The **fluctuation** model incorporates a Gaussian white noise term in the production of Nanog. If this model is true then the repopulation trajectories of the FACS sorting experiment are expected to follow rather smooth (ignoring the inherent noise) curves without overshooting. The second experimental strategy based on the single cell residence times is to yield exponentially distributed residence times, contrary to the oscillation model above.
In our FACS data we can see the smooth repopulation behavior (although not for the gates defined by the authors in [60] (LN and HN), but for LN and very high Nanog), underlining the fluctuation idea. The single cell residence times partly fosters the fluctuation assumption, too, as at least the LN cells seem to be exponentially distributed (Figure 3.18(a)), but not the HN cells.

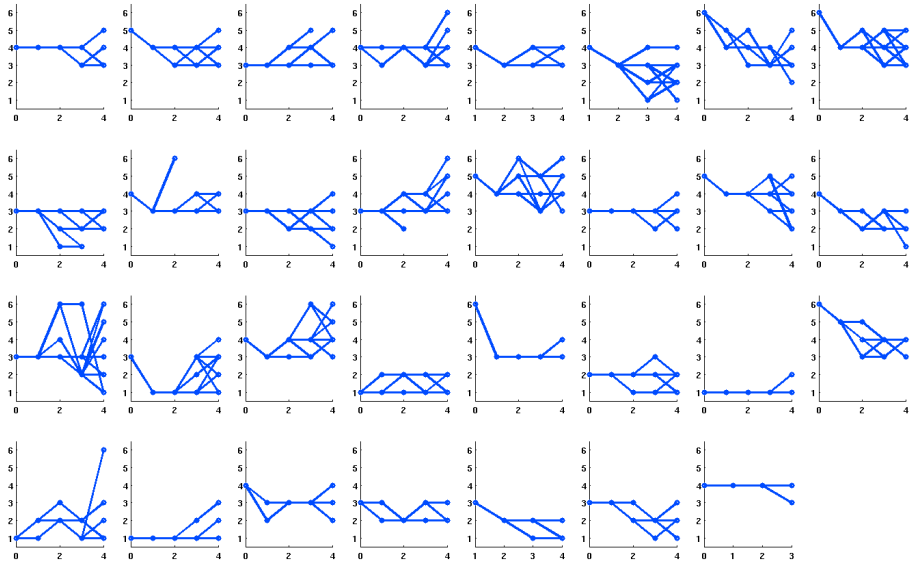
Summing up, the results of the two experimental strategies that Glauche et al. suggest to distinguish between the two model types clearly favor the fluctuation scenario, where noise impinges on the Nanog expression levels and creates the skewed distribution. However, in the analysis above we also have outlined the possible existence of more than two stable subgroups - which are not considered in the fluctuation model. Indeed, in this model an intermediate state between LN and HN cells is virtually not existent, evidenced by the section of the Nanog distribution with zero frequency and the sharp transitions between the LN and HN states, putting the bistability of the fluctuation model into question.
The discrepancy in fluctuation favoring and unfavoring data could partly be explained by the short duration of our time-lapse movies, introducing a bias in the residence time analysis, and partly by the missing definition of true time and cell cycle considerations in [60], which leaves unclear how to define the residence time in a certain state.
The reconciliation of the contradiction between our multi-stable Nanog distribution and their two stages in an extended fluctuation model would definitely be of interest as the fluctuation idea itself seems to be adequate.



(a)



(b)



(c)

Figure 3.18: Evaluation of the two Nanog models in [60]. **(a)** Residence times of cells in either the LN (left) or HN (right) state. The lines are fitted exponential distributions. **(b)** Merged schematic time courses of all trees, with the starting values of each cell sorted into either high or low. The x-axis denotes generations, the y-axis denotes the intensity compartment (high or low). The thickness of each line represents the frequency the transition occurs with. **(c)** The single time courses of all trees. X-axis denotes generations as in (b), and the y-axis the six intensity compartments from Table 3.3. To compare the six compartments with the two compartments in [60] the first two (1 & 2) represent the low state, while the rest (3 to 6) is equal to the high state. It can be seen that oscillations between the two compartments are rare.

3.4.3 A simple molecular model

As shown above, there is currently no model that can account for all experimental observations made about the fluctuating levels of Nanog in mESCs above. Two of the analyzed models seem to be either too simple or inadequate, while the third one (the fluctuation model in

[60]) possibly captures the right idea, but does not allow the existence of more than two subgroups. The compartment analysis in the previous section, in contrast, is indicative for certain subgroups, but is not a mechanistic explanation of how the distribution skewness or the subgroups are actually created. Since this might be relevant for further mESC research we develop a molecular model that is able to explain important aspects of our observed data.

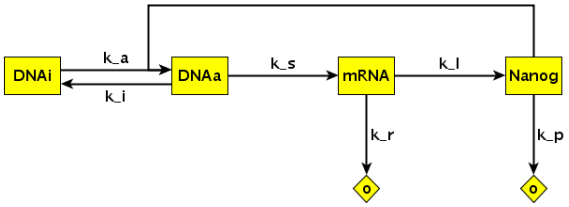


Figure 3.19: The architecture of our new molecular model. The arrows denote chemical reactions with kinetic rates k_x. DNA can be inactive (DNAi, no Nanog bound) or active (DNAa, Nanog bound at its own promoter). The transition rates between both states are k_i and k_a, respectively. mRNA is transcribed from the active DNA with rate k_s and decays with rate k_r. Nanog proteins are translated from mRNA with rate k_l and decay with rate k_p.

The model is designed as a stochastic molecular model with the architecture shown in Figure 3.19. The chemical reactions are listed in Table 3.6 and are implemented with Gillespie's direct method [57]. It contains the Nanog gene/DNA in active (Nanog bound, "DNAa") and inactive (no Nanog bound, "DNAi") mode, the Nanog mRNA ("mRNA") and the Nanog protein ("Nanog"). The self-activating role of Nanog is incorporated since only active DNA can be transcribed into mRNA. mRNA and proteins are produced and decayed during the cell cycle. The cell cycle time is sampled from the distribution of life times obtained from the movie data, with an average of 15.85 hours and a standard deviation of 4.36 hours (via a kernel density estimation). A distinction of life times according to current Nanog levels is not made, which is a reasonable simplification in the light of our analysis above (see Figure 3.14. A division exactly halves the amount of mRNA, but the proteins are segregated asymmetrically. This is incorporated by adding random numbers from a shifted and scaled t-distribution (with $df = 1.47, \mu = 0.01, \sigma = 0.12$) to the perfect division factor of 0.5 for the Nanog protein. The inclusion of asymmetric divisions is reasonable as it is observable and also important ([73]). However, random births and deaths of mRNA and protein are also known to exist, so stochastic production of mRNA and protein are also considered (inherently in the Gillespie algorithm). The loss of fluorescence is similarly handled by reducing the remaining amount of Nanog by a t-distributed (with $df = 1.65, \mu = 0.08, \sigma = 0.09$) random number. Although it is known that Nanog acts as a homodimer [119], this is not included in the model for the sake of simplicity and the implicit assumption that this inclusion would not alter the transcription dynamics dramatically.
Note that the growing cell volume over time is only implicitly regarded by the continuing Nanog transcription and translation. An example of the explicit treatment of the cell volume can be found in [37].
The measured rates for the Nanog mRNA decay k_r [153] and protein decay k_p (ISF, unpublished) are plugged into the equations. In [144] a method is proposed how to measure protein production rates and how to correct for the cell cycle-dependent increase of the production rate. Following this approach for the mESC cells with regard to the mean comparisons of

image, FACS and Western blot intensities (see above), on average 12,183 Nanog molecules per hour, are produced (7,071 in freshly divided cells and 17,296 shortly before division). However, the approach in [144] was applied to bacterial cells, leaving doubt about the applicability in mESCs, and the mean relation factor is a very coarse measure. Additionally, the production rate does not distinguish between transcription and translation, although the relative strengths of these processes are of importance [90, 130, 138, 165]. Thus the measured Nanog production rate is not plugged into the molecular model, but is left to a computational estimation.

All missing reaction rates are estimated with an optimization algorithm minimizing the cost function $f(s) = \sum_i (d(x_i) - r_s(x_i))^2$, where s denotes the current parameter state, d the reference Nanog distribution (calculated from all cells in all trees, see Figure 3.7(a)), r_s the resulting Nanog distribution for state s, and i runs over all x_i from a kernel density estimation of d. Both distributions are normalized to an intensity range of $[0, 1]$. A perfect match between the simulated and the reference Nanog distribution would result in $f = 0$, and the quality of a state/fit s is measured by f in the following (the lower, the better the fit).

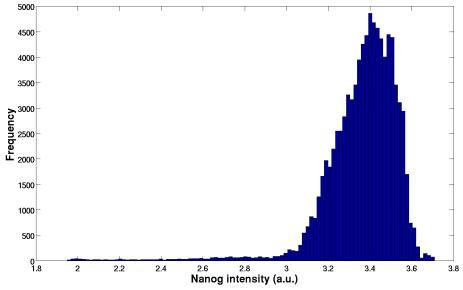
The optimization is done by a Simulated Annealing (SA) algorithm [87] which is run several times with different starting parameters. One of the start value vectors has been selected manually, estimated from a non-exhaustive manual search space inspection (see column 2 of Table 3.6), the others are random. All SA runs converge into (approximately) the same parameters, given in column 3 of the same table.

Name	Chemical reaction	Measured rates (h^{-1})	Estimated rates (h^{-1})
R_1	DNAi + Nanog $\xrightarrow{k_a}$ DNAa	[0.1]	0.0847
R_2	DNAa $\xrightarrow{k_i}$ DNAi + Nanog	[0.01]	0.0087
R_3	DNAa $\xrightarrow{k_s}$ DNAa + mRNA	[6]	3.4284
R_4	mRNA $\xrightarrow{k_l}$ mRNA + Nanog	[15]	28.6904
R_5	mRNA $\xrightarrow{k_r} \emptyset$	0.12	
R_6	Nanog $\xrightarrow{k_p} \emptyset$	0.193	

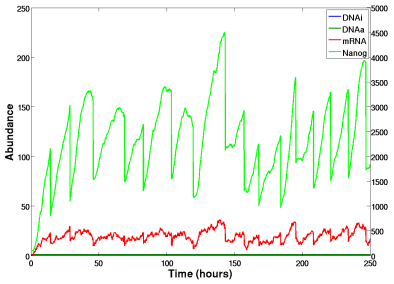
Table 3.6: Reactions and rates in our stochastic molecular model. DNAi and DNAa denote inactive (no Nanog bound) and active DNA (Nanog bound), mRNA denotes Nanog mRNA and Nanog denotes the Nanog protein. The decay rates for the mRNA k_r and protein k_p are experimentally measured (second column), while the other reaction rates are estimated by an optimization algorithm (third column). The rates in square brackets in column 2 denote the manually selected starting values for the optimization, which has been obtained from a manual search space inspection.

3.4.4 Validation of the new model

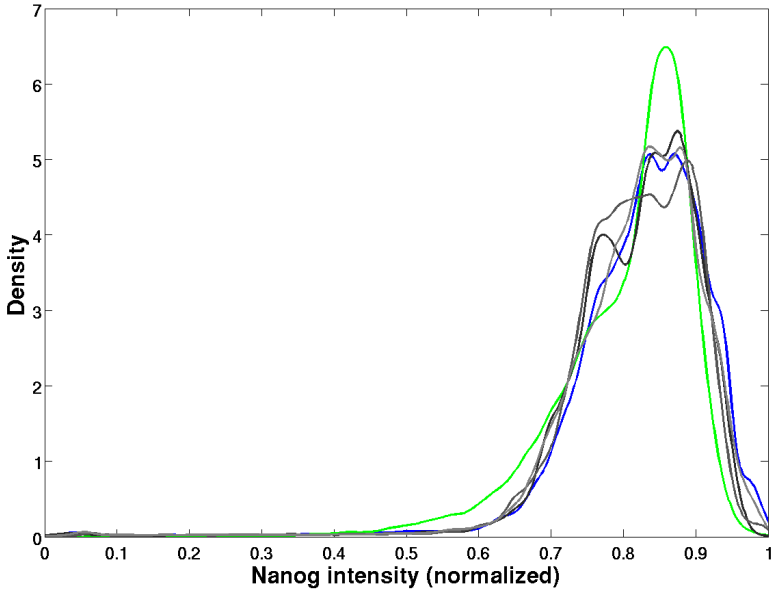
After the parameters of the model have been estimated, a validation of the model is performed to test its adequacy in the light of our data. There are five observed attributes of the single ESCs or their Nanog intensity distribution that the model should explain: (i) the shape of the Nanog distribution, (ii) the steadiness of this distribution, (iii) the existence of a distinct number of stable subgroups, (iv) the Nanog expression memory of cells and (v) the repopulation of the steady state distribution of Nanog.



(a)



(b)



(c)

Figure 3.20: Gillespie simulation results. **(a)** The Nanog expression distribution obtained from the simulation, started with a random abundance of Nanog taken from the original Nanog distribution. The rate parameters of the reactions are provided in Table 3.6, and are estimated to create a close match between this distribution and the reference distribution from the single cell movies in Figure 3.7(a). The x-axis shows the Nanog intensity (in arbitrary units, log-scale), the y-axis their frequency. **(b)** An example time course of the Gillespie algorithm. Each line shows the trajectory of one species. The x-axis shows the time in hours, the y-axis the species abundance (left y-axis is for DNAi, DNAa and mRNA, right y-axis is for Nanog). The sharp decays in mRNA and Nanog abundance are the divisions. **(c)** Steadiness of the simulated Nanog distribution. The blue line is the same distribution as in (a), the green line is the movie distribution from Figure 3.7(a) and the gray lines are simulations with random start values for the Nanog abundance sampled from the blue distribution. It can be seen that the resulting distributions are similar to the original one, indicating its steadiness.

The first point (i) is obviously fulfilled, as the parameters are selected to yield exactly such a distribution. But one can ask which of the model components (asymmetric division, fluorescence loss during division, and different cell life times, half-lifes of mRNA and protein)

are actually responsible for generating this skewed shape. To assess this each of the four components is switched off in turn, and the quality of the fit is calculated for the resulting model. The fit quality of the best solution is $f(s) = 63.5$, with s the four parameters in Table 3.6. The resulting fit qualities of the impaired models are given in Table 3.7. The component with the highest impact on the fit quality is the asymmetric division, suggesting that this feature is responsible for a large part of the heterogeneity in Nanog expression. The divisional fluorescence loss has the least impact, indicating low relevance for the distribution skewness. The equality of life times (division at a fixed time of 15.85 hours) and the swap of the protein and mRNA half-lifes have roughly equal and rather high effects, showing their importance in the process.
To test the high influence of the division asymmetry experimentally, one could perhaps plate one ESC on a culture dish with a surface structure (e.g dips and peaks) that forces the cell to divide asymmetrically, and an ensuing inspection of the Nanog distribution in both daughter cells over several generations.
A parameter sensitivity assessment is performed as well. Each of the six reaction rates is altered between 20% and 200% of its original value while the others are left constant. The impact on the fit quality of each change is given in Figure 3.21. The variation of each parameter corresponds to one subplot. The x-axis shows the relative value and the y-axis the quality fold change in comparison with the optimum quality. It can be seen that, as expected, the estimated parameters have the best quality. The larger the relative change, the worse the correspondence with the reference Nanog distribution in general, with the exception of k_a. The parameter with the least impact is the DNA inactivation rate k_i, while the highest impact can be observed if the transcription rate k_s is strongly reduced. The two parameters with experimentally measured values, k_r and k_p, show similar effects, but with the interesting exception of even further reducing the quality if the mRNA decay is slowed down or the protein decay fastened.
Altogether, the analysis underlines the optimality of the estimated parameters and indicates the necessity of the cell to keep the reaction rates within a certain range. An experimental strategy to assess the validity of this sensitivity measurement could consist of siRNA introduction (speeds up mRNA decay and slows translation), Nanog ubiquitination enhancement (speeds up protein decay), the suppression of Nanog binding at its own promoter (reduces k_a) or the constant activation of this promoter (enlarges k_s).

Model	Fit quality	Fold increase
Original	63.5	1
No asymmetric division	291.2	4.59
No loss in division	130.38	2.05
Equal life times	259.6	4.09
Half-lifes exchanged	239.44	3.77

Table 3.7: Changes in the simulated Nanog distribution after separate model destructions. The first column denotes the changed component, the second the resulting fit quality (higher values denote a worse fit). The third column is the fold increase of this quality over the original one given in the first row.

Point (ii) about the distribution steadiness can be tested by starting the simulation with different starting values taken from this distribution and checking if the complete distribution

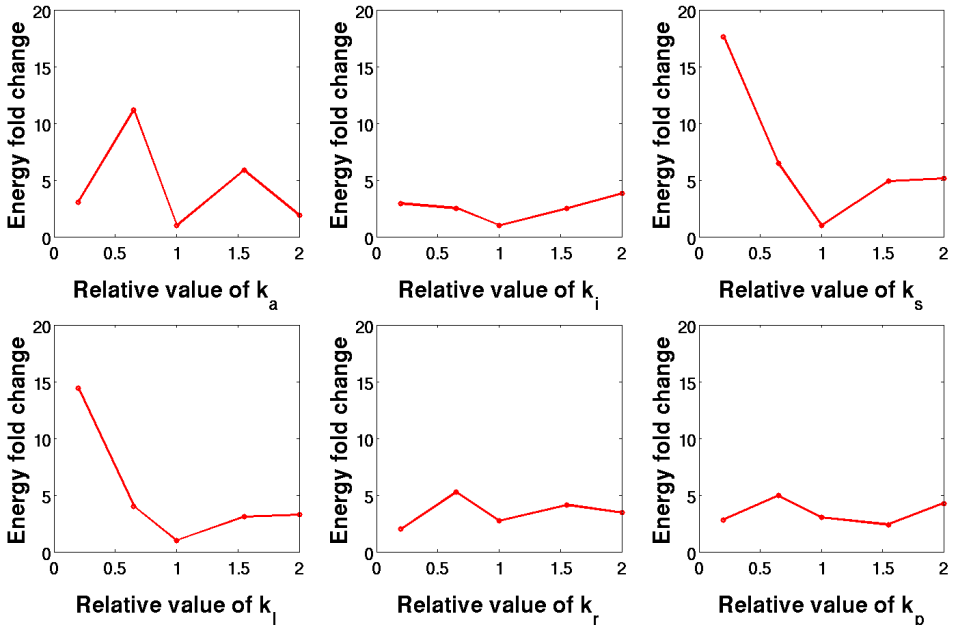
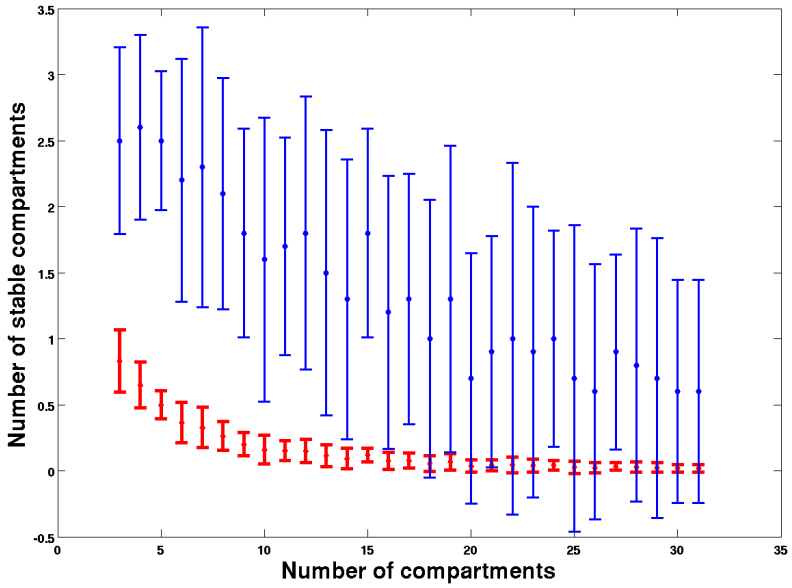


Figure 3.21: Parameter sensitivity of our molecular model. Each subplot shows the sensitivity analysis for one of the six reaction rates in Table 3.6. The x-axis denotes the relative parameter value (from 20% to 200%), the y-axis the quality fold change with respect to the best (minimum) fit quality.

is obtained again at the end. This is shown in Figure 3.20(c), where the resulting distributions of ten simulations with different starting values are plotted (in shades of gray) and compared to the original simulated one (in blue) and the distribution of the movies (in green). The shapes and locations of the gray and blue curves are similar, and their close relationship is additionally evidenced by the very low KL divergence of 0.02 (on average) of the gray curves to the original simulated distribution (blue).

Point (iii) is best addressed by subdividing the Nanog start values of the Gillespie simulations into different compartments and measuring the transition rates between these. Similar to above, the distribution is separated into equally spaced quantiles and the number of stable compartments (defined as $P_{xx} > 0$) counted over several runs. The results are shown in Figure 3.22(a). The plot suggests the existence of several subgroups, independent from the actual number of compartments. The error bars denote the standard deviations. Although these indicate substantial variations, the general trend (also in the fraction) is obviously to have less stable compartments the more compartmentalized the distribution is. In the case of three or four subgroups, most of them are stable. Altogether, we state that the existence of subgroups is a good assumption.

The memory property (iv), which states that the future Nanog expression of cells is dependent on their history is addressed equally as above, with calculating the transition rates between different compartments when regarding the cells' history. If there are differences, the history of expression is indeed of importance. The distribution is split into three subgroups (a conclusion from the previous analysis), and the change rates are provided in Table 3.8, in a



(a)

Figure 3.22: The existence of stable compartments or subgroups in the new molecular model. The x-axis shows the number of equally spaced compartments in the distribution, the y-axis the number (blue points) and fraction (red points) of stable compartments (with $P_{xx} > 0$) similar to Figure 3.15. The stable compartment count and fraction are averaged over multiple runs, and the error bars denote the standard deviation.

way similar to Table 3.5. It can be seen that indeed the change rates differ, depending on the provenience of the cell, which would point to an expression memory. Although this memory effect depends on stochasticity only and the stability of these numbers is not assessed here, it could be concluded that cells have an inherent memory. However, a similar effect as above (Table 3.5), where cells indeed seem to heed their history is not seen here.

Property (v) addressing the repopulation of the steady state levels is already evidenced in the steadiness of the distribution, where after several generations the original Nanog distribution is restored independent of the starting value. Additionally, the steady state concentrations of both mRNA and Nanog can be calculated analytically if their production and decay are modeled as ODEs (ordinary differential equations, see also chapter 2). DNA (in)activation is neglected, which is reasonable due to the small reaction proportions in comparison with the others and is a simplification often applied [150]. The equations in Figure 3.23 show how to derive the steady state concentrations, where m is the amount of Nanog mRNA, N the amount of Nanog protein, a dot above a symbol means the derivative over time (the change of concentration over time) and the values of the reaction rates k_x are taken from Table 3.6. To be a steady state, the derivative must be 0 as there is no change in the concentration by definition. This steady state is reached independently of the starting concentrations - they do not occur at all in the equations. Thus the existence of a steady state and the ensuing reestablishment of it is shown mathematically, too.

		Daughter cells		
Mother	Current	1	2	3
1	1	0.33	0.33	0.33
2		0.11	0.33	0.56
3		0.43	0.57	0.00
1	2	0.40	0.40	0.20
2		0.67	0.17	0.17
3		0.17	0.17	0.67
1	3	0.30	0.60	0.10
2		0.30	0.30	0.40
3		0.14	0.43	0.43

Table 3.8: Change rates between compartments in the simulated Nanog distribution, regarding the ultimate cell history. The rows within one current compartment are different, indicating an influence of the history.

$$
\begin{aligned}
\dot{m} &= -k_r * m + k_s \\
\dot{N} &= -k_p * N + k_l * m \\
\dot{m} &\stackrel{!}{=} 0 \Leftrightarrow m = \frac{k_s}{k_r} = 28.58 \\
\dot{N} &\stackrel{!}{=} 0 \Leftrightarrow N = \frac{k_l}{k_p} * \frac{k_s}{k_r} = 4248.50
\end{aligned}
$$

Figure 3.23: Calculation of the steady state Nanog mRNA (m) and protein (N) levels with the help of ODEs.

Interestingly, we can observe an average 2.23 (sd = 0.2) fold increase of Nanog intensity during the cell cycle in the simulations. The fold increase is measured by dividing the last intensity value (mean of the last two hours) by the first intensity value (mean of the first two hours) of each cell and averaging over all cells. This corresponds well with the expected value of 2, as the amount of protein must double during the cell cycle if the levels after division should be equal to the mother cell level if there is no loss during division. This is often the case in the real cells, evidenced by the stability of the compartments and the change rates (Table 3.4). Although the increase factor measured in the time-lapse movies is higher (2.58, sd = 1.81), one could state that the factors are comparable and thus prove the adequacy of our model idea. Additionally, the small standard deviation in the simulation fosters the existence of distinct subgroups since it ensures rather equality than difference between mother and daughter cell Nanog levels.

Summing up the results of our new molecular model, we can state that both the skewness of the Nanog distribution and the existence of interchanging, but stable subgroups is possibly a result of simple mechanistic processes in the cell. The most relevant one seems to be the asymmetric segregation of Nanog molecules during division, evidenced by the high quality fold change if it is switched off. This asymmetry can result in a long-term influence on future Nanog levels as shown in [73].

3.5 Discussion

In this chapter we have studied the expression pattern of Nanog, a protein with vital role in pluripotent mESCs. Two types of experiments have been described and the resulting data have been analyzed. Existing theoretical models of the fluctuations in expression have been compared with these data and a new molecular model has been proposed. In general, the power of single cell imaging over population studies has been underlined.

3.5.1 Result summary

In the first part a brightness normalization of the time-lapse image intensities based on fluorescent beads has been introduced and applied to our movie data. A comparison of the fluorescence intensities obtained from a FACS population analysis and the microscope images has been performed, indicating the correlation between both methods to quantify fluorescence. Thus we conclude that the time-lapse imaging approach is applicable in general and allows for the generation of highly resolved single cell data.

By an accurate quantification we have shown that the cell division is not perfect. First, the amounts of Nanog distributed on the nascent daughter cells are not equal, but rather follow the right half of a shifted and scaled t-distribution. Second, a substantial amount (on average 8%) of the Nanog protein is lost during division (assuming that the loss is not only due to technical effects but a real feature). There are several possible reasons for this effect: the cell could either actively or passively transport Nanog from the nucleus into the cytoplasm during mitosis, or Nanog is degraded during division, or a rapid shape change before and after division (observed in some cells, data not shown) creates a quantification artifact, or the quantification is inaccurate. Assuming grave measurement problems to be unlikely (evidenced by the correspondence between single cell and population distribution), the other three possibilities remain. To test them experimentally a quantification of the cytoplasm would be helpful because currently only the nucleus is quantified. Additionally, a shorter time interval between taking two pictures could bring clarification since a division lasts for only about 20 minutes while images are taken every 30 minutes. Current cytoplasm quantification experiments in the ISF laboratory seem to foster both the transportation and degradation or shape change hypothesis (data not shown). The experimental strategy suggested in [73] to distinguish between asymmetric partitioning or stochastic protein production as the source of heterogeneity could be applied here, too. It could also result in differences to the current results if the division characteristics of non-adherent cells are ignored, as these are not stuck to the culture dish and can roam across the surface while changing their shapes, resulting in possibly inaccurate measurements.

We have surveyed the reestablishment of the steady-state skewed Nanog intensity distribution from sorted cells both in mESC populations and single cells. The results of both methods are not completely correlated, which could primarily be a consequence of FACS measurement errors or the short duration of the movies. But another explanation is possible, which is more intriguing: the lowest 2% of cells (which were sorted in FACS) are truly different from the lowest 10% (which have been selected in the movies), comprising differences in repopulation speed and capacity. However, the general trend seems to be equal in both experiments and conveys that even cells that are taken from the extremes of the Nanog intensity distribution are capable of fully regenerating the whole population of cells with various Nanog expression levels (shown in FACS by the six week experiment). This is especially true for cells with a

medium Nanog expression that almost perfectly reestablish the complete intensity spectrum within three rounds of division. For the cells from both extremes this takes more generations, since they have to roam across the full expression space. Altogether, the full Nanog distribution seems to be both a steady state (as it is observed for any random sample of mESCs) and an attractor (as it is regenerated over time). These results are consistent with a non-genetic memory which is passed from generation to generation, but with decreasing correlation and strength [1, 83, 154]. This finding is also confirmed by measuring the influence of the cells' history on their future behavior, clearly indicating that previous expression levels are remembered, but are gradually lost over time. Longer lasting quantitative studies in single cells (more than four generations) could foster these insights and additionally give hints how this expression memory is maintained or degraded. Specifically inspecting very low cells (2%) in the movies or sorting only the lowest 10% in FACS could help to reconciliate the two types of experiments. A theoretical extension (if sufficient data is provided) could be the fitting of a multi-stage Markov process to the data, with a history of one or two generations, and testing it against the observed repopulation.

The cell cycle was not regarded in former studies of the fluctuating Nanog levels, neither theoretically nor experimentally, and the question has been asked whether the cell cycle-dependent increase of Nanog and its asymmetric halving in division is responsible for the skewness of the intensity distribution. By taking only the measured intensities right at the start of the cell cycle, we have shown that the form of the distribution is truly regulated by the cell. Although the Nanog production during the life time of cells slightly shifts the distribution to higher values, it does not change its general shape. As a consequence we can state that the production and degradation during the cell cycle is not important for regulating heterogeneous protein levels. This result cannot be obtained from a population analysis but only from single cell data, which strengthens the importance of single cell analysis as in the population data valuable information could be masked.
An experiment with synchronized ESCs (where all cells are roughly in the same state of the cell cycle) or a time-lapse movie with cell cycle markers like PI (propidium iodide) or PSLD (phosphorylation-dependent subcellular localization domain) could help to validate our conclusion.

Detailed measurement of the cell cycle times revealed that even ESCs with very low Nanog expression have equal proliferation rates to those with very high levels. This is intriguing as it could be argued that high Nanog-ESCs are highly pluripotent, whereas low Nanog-ESCs are poised for response to differentiation signals and thus would exhibit different life times – ESCs proliferate rather rapidly, whereas differentiated cells have slow cycles or do not proliferate at all. As a conjecture we state that the life times of ESCs are not apt to distinguish between different levels of differentiation commitment.

Our separation of the Nanog intensity distribution into compartments revealed the existence of various subgroups with distinctive levels of Nanog expression. This is conform with former identifications of subgroups in various protein levels [66, 174]. We have shown the stability of these subgroups by the substantial propensity of cells to stay within their compartments. The existence of subgroups could be validated experimentally with an ESC line that bears fluorescent markers for both Oct4, a pluripotency marker, and Nanog. If the cells expressing Oct4 and a certain level of Nanog can either keep this level of Nanog (within a certain range) or convert into another group, these cells could be considered a true subgroup (similar to [174]). A similar line (with Sox2 instead of Oct4) is currently under construction in the ISF laboratory.

The behavior of individual mESCs can be summarized in one sentence: they are flexible in Nanog expression, but prefer higher values, and eventually tend to establish a broad steady-state spectrum of Nanog expression independent of their original expression level, although they only slowly forget their origin.

Having gained valuable insights into the Nanog intensity distribution in single cells, the question about the mechanistic origin of this distribution has been treated in the modeling section. In the beginning, three existing molecular models have been validated against our data. It has turned out that two of them are opposed to our data, raising doubt about their adequacy. The third model, based on random unspecified fluctuations in Nanog production, corresponds with our population and single cell data in some aspects, but not all. Although an extension of this model would be of interest, too, a new molecular model has been proposed.
This model includes the measured characteristics of the ESCs taken from the single cell movies and previous studies. These are the asymmetric segregation and loss/degradation of Nanog molecules during division, the distribution of cell life times and the decay rates of both Nanog mRNA and protein.
Since several observations from the movies can be observed in our model, too, we state that it is highly recommendable to try simple approaches first before delving into heavily parameterized models that need to rely on a manifold of unmeasured factors. Indeed, it would be very interesting to measure experimentally the decay rates (mRNA and protein) of other fluctuating ESC genes like Stella or Rex1. Additionally, we have seen that the odd half-life ratio of Nanog mRNA and protein (normally, the mRNA decays much faster than its protein) can provide a transcriptional memory and could be responsible for some of the features observed in the single cell movies. Many models of gene expression assume that the half-life of mRNA is (much) smaller than that of proteins, making it possible to derive analytical solutions of the simple DNA-RNA-Protein model [134, 150, 165, 172]. This is not the case for Nanog and possibly some other genes, too, disallowing the neglection of transcription over translation. Intriguingly, a recent publication derives mathematically that the segregation asymmetry during division dominates all other noise sources, if the half-life of protein or mRNA is on the scale of the cell cycle (which is the case for Nanog) [73]. These two facts create a necessity for new models of gene expression that take these facts into account.
It would be interesting to see the resulting distribution of the same Gillespie simulation if the half-lifes of Nanog mRNA and protein were exchanged with the ones of eGFP. If the distribution is bimodal this would be a clear indication for the necessity of a fusion line instead of a reporter line, as it is more accurate (the bimodality would then be due only to the different half-lifes of GFP and Nanog, and not a feature of the Nanog expression itself).
However, our model is not able to explain the more complicated properties of the ESCs like the memory, and it would definitely be of interest to continue searching adequate mathematical representations of these features. Additionally, the reaction rates are measured with a SA approach only, which provides us with the maximum likelihood estimators, but not the parameter distribution. Thus a further step could be the error estimation of these parameters (e.g. by bootstrapping), and a Bayesian approach or Markov Chain Monte Carlo (MCMC) algorithm to derive the distributions of these rates would be interesting.
Since the model we have suggested is rather simple and its only regulatory element is a positive feedback loop, we suggest that the heterogeneous expression of one gene is actually the

default state and genes that are expressed within a narrow range (e.g. Oct4) need tighter or more intricate regulation. It needs to be stated that heterogeneity does not necessarily coincide with fluctuations, *i.e.* interchangeable levels of expression, which might truly be a feature of a few genes only [71]. Recently, a study on the fundamental limits of noise suppression [99] showed that stochastic processes in the cell, be they desired heterogeneity or just nuisance, cannot be neglected.

3.5.2 Outlook

Even though some of the secrets of ESCs with respect to Nanog expression have been revealed with the detailed and quantitative analysis of single cell time-lapse movies, the cells keep their incredible and scarcely understood complexity. Additionally, all the observations discussed here are made in *in vitro* experiments, which leave unclear the applicability and validity *in vivo*. To cure this problem, *in vivo* imaging is an appreciable tool that could generate novel and unprecedented insights. Such systems are currently under development, but not yet applicable to whole organisms [148].
The medical importance of the Nanog fluctuations are currently not assessed, but fluctuations in general can have important effects, e.g. in the treatment of cancer (where some heterogeneous cells are able to resist drug treatment) or the reprogramming of differentiated cells to iPSCs [189].
The stochastic transcriptional and translational rates used in our model are actually just a summary of more complicated effects like ribosome count, mRNA or protein elongation time, the number of available tRNAs and numerous others, which all could be considered in a more detailed molecular model of the Nanog expression. Furthermore, the exact values of the reaction rates do not depend on Nanog alone, but are rather the result of thousands of other genes with distinct expression levels that contribute to shaping these factors, which thus are prone to change over time. Therefore, the extension of this simple molecular model by further players could be of interest. A study with highly interesting single cell resolved data for such a model extension is by Guo et al. [62] (already used in chapter 2), which provides expression and regulation patterns for a set of genes associated with pluripotency and development. Another possible direction is the combination of the fluctuation model by Glauche et al. [60] with a memory module to account for the existence of more than two subgroups and more detailed sources of the fluctuations they utilize.
An attempt to decipher the absolute abundance of mRNA in ESCs from a previous study [24] has failed due to inaccuracies in its measurements and unclear calculations, but the coarse results suggest a very low copy number of mRNA (about ten per cell on average) in comparison with the amount of protein (about five magnitudes higher). Whether the prolonged half-life of the Nanog mRNA in comparison is a consequence of this low abundance or the driving force behind it is an interesting question to be resolved.
The results for our analyses are obtained from single cell imaging, which is able to generate population level data by simple aggregation. This is not possible the other way round as population data (like FACS) cannot be resolved to single cells. Concluding from the slight discrepancies between the FACS population and the single cell data and the impossibility of obtaining the cell cycle characteristics or memory effects from aggregated population data, we state that single cell analyses only can help to shed light on these issues.
A transfer of our results and models to hESCs cannot be assessed because of lacking Nanog expression data in these cells. This would be worth future investigations, since the ultimate

goal of stem cell research is the understanding of human diseases and their elimination. On the long path ahead of us, the first tentative steps are already made.

Chapter 4

Outlook

In the previous two chapters, two different levels of gene regulation and interaction have been studied. First, a Boolean network has been constructed to explain the properties and differentiation decisions of ESC under certain culture conditions on a systemic perspective. Second, the expression of a single gene, Nanog, has been analyzed from single cell time-lapse movies and several characteristic features of this gene have been extracted. Additionally, previous theoretical models of Nanog expression have been evaluated and a new molecular model has been proposed. Biological hypotheses have been developed from both approaches and experimental strategies have been suggested to test them.
Although a Boolean network might be considered far too simplistic and even the molecular model too coarse-grained, the correct level of abstraction is hard to define [14, 105] since each stage has its own advantages and drawbacks and is fruitful for studying certain aspects and answering a defined set of questions. Of course, understanding the details of each gene regulation is of outstanding importance since it ultimately drives cell behavior and development and eventually life itself. However, often enough an integrative view is sufficient to understand cellular patterns or components and mercifully allows the experimentator to disregard the chemical details.
One possible direction for further research is the modeling of cell populations, similar to [88]. Cells could be represented by agents with either a deterministic or stochastic internal program, responding to external stimuli on the one hand, and excreting cell-cell communication factors on the other hand. This model could account for the cross-talk between cells in the developing embryo [98] and include spatial or polarity constraints that drive cell fates [20]. A similar approach has already been examined in [69], where the minimal requirements for blastocyst formation are analysed with computational simulations.
This work is only a further tessera in the large picture of mammalian embryogenesis. There are countless levels of intermediate complexity between the two studied here, important without doubt and far from being understood. The clinical potential of ESCs cannot be estimated high enough currently, and future research will surely reveal further insights paving the way from bench to bedside.

Chapter 5

Appendix

5.1 Reference list for the embryonic development tree

In Table 5.1 the references (letter + number) used in the embryonic development tree (Figure 1.2) are mapped to the literature in the bibliography. The entries are ordered alphabetically.

Key	Article	Key	Article	Key	Article	Key	Article
A3	[5]	A9	[2]	B0	[19]	B6	[16]
B8	[17]	C8	[22]	E9	[47]	F2	[54]
G0	[62]	H0	[64]	I0	[75]	I6	[74]
K0	[93]	K5	[84]	L0	[96]	L6	[95]
M0a	[115]	M0b	[108]	M8	[106]	N0	[125]
N7	[126]	N9	[123]	O9	[129]	P7	[137]
S5	[159]	S6	[163]	S9	[146]	V0	[177]
V9	[176]	W8	[180]	X9	[188]	Y5	[191]
Y9	[194]	Z0	[195]				

Table 5.1: Reference keys used in the embryonic development tree (Figure 1.2) with the corresponding articles.

5.2 Proteins and interactions in the Boolean networks

Each gene/protein used in the two Boolean networks CME and NCC in chapter 2 is found in the PluriNetWork, indicating a role in either the maintenance or loss of pluripotency. The following short list provides an overview of their roles in murine embryonic stem cells, condensing the manifold information available for each of them.

- Oct4/Pou5f1 (OCTamer-binding transcription factor 4, or POU-domain class 5, TF 1) is known as a marker of mESCs, indicating the loss of pluripotency when down-regulated [23, 122].
- Sox2 (Sry-bOX containing gene 2) plays a role in both maintaining pluripotency [5] as well as neuronal differentiation of cells [51]. Sox2 forms a heterodimer with Oct4 that

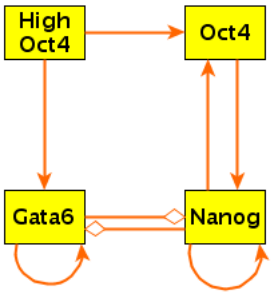
triggers its own reciprocal activation [32] and the activation of several other pluripotency genes [101].

- Nanog is a fundamental actor in mESCs [26, 114], also called the 'safeguard' of pluripotency [27]. When Nanog levels are high, ESCs can be maintained even in the absence of LIF [26], while cells with low Nanog levels are possibly open for differentiation [27].
- Cdx2 (CauDal-type homeoboX 2) is a driver of TE formation and an antagonist of Oct4 in the decision between TE and ICM [128].
- Gata6 (GATA-binding protein 6) is required [22] and sufficient [54] for the formation of PE. It is also expressed in the TE [62].
- Gcnf/Nr6a1 (Germ Cell Nuclear Factor or Nuclear Receptor Subfamily 6, group A, member 1) is a repressor of Oct4 and Sox2 [127] and Nanog [117] and is thus essential for silencing the pluripotency machinery after or during differentiation.
- Fgf4 (Fibroblast Growth Factor 4) is activated by the pluripotency triad Nanog, Oct4, Sox2 and down-regulates their expression via the FGF-signaling pathway [18, 96]. This suppression is mediated by Erk and marks the onset of differentiation.
- Erk (Extracellular signal-Regulated Kinase) suppresses the pluripotency genes (especially Nanog, Oct4 and Sox2) and its expression is a requirement for differentiation of ESCs into EpiSCs or further states [18, 96].
- Rex1 is a distinction marker of mESC (expressed) vs. mEpiSC (not expressed) [18]. Although this view is currently under revision [64], the expression of this protein is used to distinguish between the two different pluripotent states.
- LIF (Leukemia Inhibitory Factor) is used together with a specific serum (containing nutrients and other factors, like BMP4) to maintain mESCs *in vitro* for unlimited time and passages [161].
- Activin is used together with FGF2 (Fibroblast Growth Factor 2) to arrest mEpiSCs in their current state [18, 42, 176] for indefinite time *in vitro*.

5.3 A tristate Boolean motif

In [33] a network motif consisting of Nanog, Oct4 and Gata6 is proposed. It tries to model both the activation of Nanog by Oct4 at intermediate levels and its suppression at high Oct4 levels. The activation is direct but the suppression is effected by Gata6 as an intermediate player. This motif is implemented with ODEs in this paper. In chapter 2 of this thesis the suppressive effect of high Oct4 levels is not treated, reasoned by the necessity of three distinct states (low, medium and high Oct4) in a Boolean network, which allows only two states by definition. However, the implementation of a tri-state node is not impossible, evidenced in Figure 5.1. The image shows the connection pattern of Nanog, Oct4 and Gata6 as in the CME network in chapter 2, but with an additional "High Oct4" node. The latter node is on (1) if the levels of Oct4 are high, but off (0) if the levels are low or medium. The table shows the states/expression values of Nanog and Gata6 under different Oct4 levels. The first row represents low or absent Oct4, the second medium levels and the last one high Oct4 levels. It can be seen that an equally graded effect of Oct4 levels on the expression of Nanog can be achieved similar to the ODEs, but without any required parameters. The idea of the motif is

based on [67].
The reason why this motif is not included in the analysis above is the lack of connections and activation patterns between the different nodes – when is "High Oct4" on, and when is it off? Additionally, to our knowledge there is no literature evidence of such a wiring as proposed in [33]. Nonetheless, this motif shows the theoretical capacity of Boolean networks to capture more than two states of one node, but at the expense of an exponentially growing network size.



Oct4	High Oct4	Nanog	Gata6
0	0	0	*
1	0	1	0
1	1	0	1

Figure 5.1: A possible Boolean network motif for capturing tristate Oct4 levels and the expression levels of Nanog and Gata6 under different levels of Oct4. In the first case (no Oct4), the Gata6 level is unspecified (*).

5.4 Abbreviations

The abbreviations used in the text are listed in Table 5.2, in alphabetical order.

5.5 Fluorescence analysis

In chapter 3 a model for the distribution of Nanog in mESCs has been proposed. This model has been driven by and validated against fluorescence intensity data of Nanog generated in Timm Schroeder's laboratory[1] by two distinct methods. The first is fluorescence activated cell sorting (FACS), a common laboratory technique to sort cells according to various attributes like size, granularity or fluorescence intensity. Hierarchical filter gates can be defined to further refine the sorting. The result of the FACS process is a set of cell populations with roughly equal attributes within each population. These populations can be subjected to further analysis, for example culturing under certain conditions or imaging with a microscope.
The second methodology used to generate the data analyzed in chapter 3 makes use of a unique time-lapse microscopy setup. A picture of the workflow is provided in 3.1 in chapter 3. In the first step a population of fluorochrome-stained cells is selected, for example by FACS, and plated on a well. The second step is the long-term observation of this well with a Zeiss microscope capable of taking phase contrast (bright field) and fluorescence pictures with different excitation and emission wavelengths. A well on the plate ('experiment') can be segmented into different movie positions. The fluorescent activity of the cells is generated either by genetically modified proteins (in our case Nanog is fused with eGFP) or antibodies.

[1] http://www.helmholtz-muenchen.de/isf/haematopoese

Abbreviation	Meaning	Abbreviation	Meaning
AF	Adam Filipcyzk	BMP	Bone morphogenetic protein
CME	Chickarmane Model Extended	Epi	Epiblast
FACS	Fluorescence Activated Cell Sorting	FGF	Fibroblast Growth Factor
GFP	Green Fluorescent Protein	GRN	Gene Regulatory Network
HN	High Nanog expression	HSC	Hematopoietic Stem Cell
ICM	Inner Cell Mass	iPSC	Induced Pluripotent Stem Cell
ISF	Institut für Stammzellforschung	LIF	Leukemia Inhibitory Factor (or the German license tag of Lichtenfels)
LN	Low Nanog expression	(m)EpiSC	(murine) Epiblast Stem Cell
(m)ESC	(murine) Embryonic Stem Cell	NCC	Network with Culture Conditions
NSC	Neuronal Stem Cell	(O)DE	(Ordinary) Differential Equation
PE	Primitive Endoderm	PTM	post-translational modifications
TE	Trophectoderm	TF	Transcription Factor
TTT	Timm's Tracking Tool		

Table 5.2: Abbreviations used in the text, ordered alphabetically.

Imaging can be sustained up to a duration of weeks, taking pictures at defined intervals usually in the minute or hour range. One such experiment produces up to a terabyte of image data, stored on mobile hard disks. In step three the pictures are analysed with a software suite called TTT (Timm's Tracking Tool). TTT is designed to facilitate the persecution of single cells on the images, enabling the creation of cell pedigrees in a highly time-resolved manner. The attributes of single cells like size, motility, morphology or fluorescence intensity, and its decisions, namely division, differentiation or commitment to a final fate can be surveyed and comfortably annotated by continuously tracing these cells without losing their identity. This technique allows for answers to important questions that had been unresolved for decades due to the lack of such continuous single-cell analysis tools (see [44, 142] for examples). TTT greatly enhances the experiment throughput by providing a reliable, scrupulously maintained framework that allows the user to focus on the main aspects of his project and by automatising commonly used procedures like intensity quantification and statistical analyses. The author of this thesis has been working on TTT software for more than five years; he has designed the programming architecture, implemented all tools in tune with the experimentators' needs and maintained the software usable over this whole period. A screenshot of the TTT program is shown in Figure 5.2.

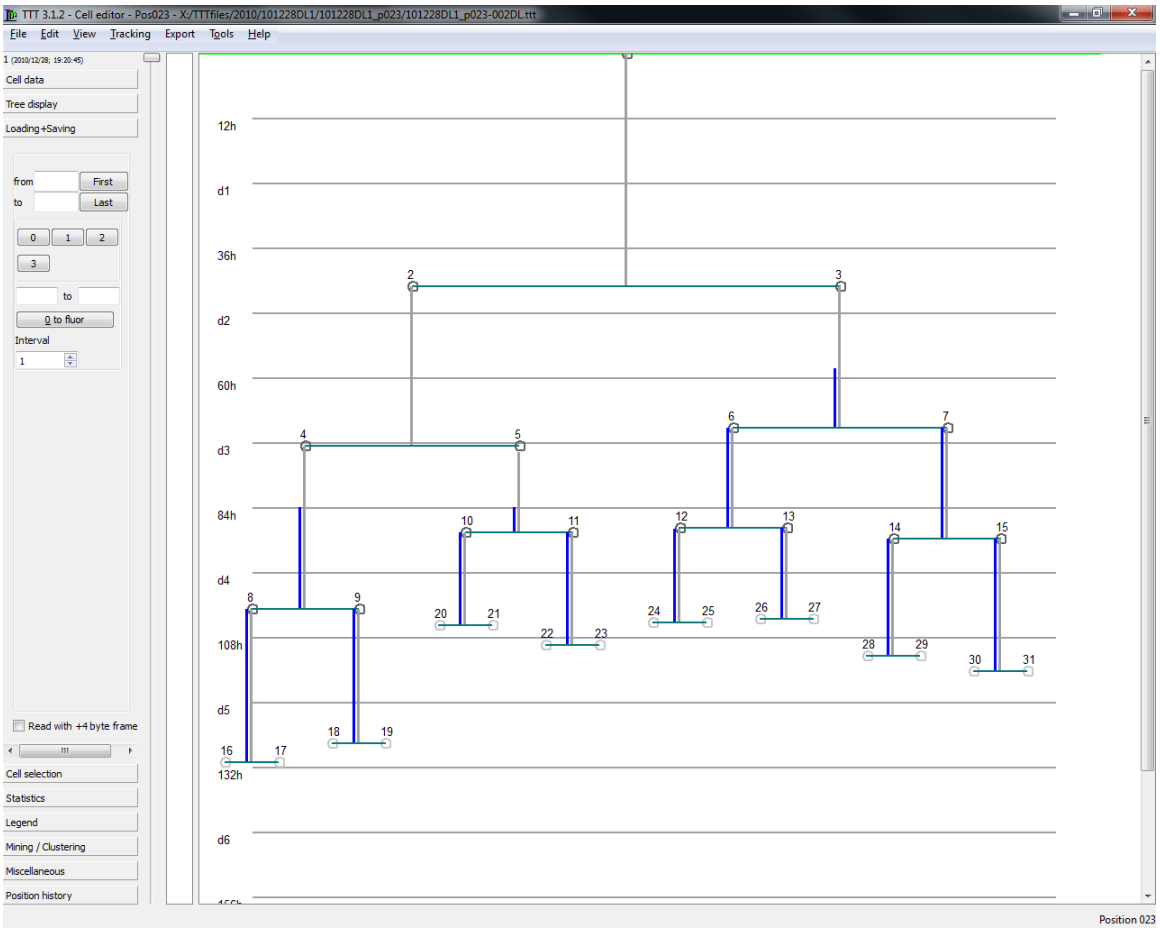


Figure 5.2: Screenshot of TTT, the program used to generate the single-cell genealogies from the time-lapse images. It allows comfortable cell tracing, feature annotation and data analysis in one pour.

5.6 Software

All calculations were performed in **Matlab**, Release R2009b, provided by The MathWorks[2]. Graphs and networks were drawn in **yEd**, Version 3.6, provided by yWorks[3].
The Boolean network analysis was performed with **Odefy**, kindly provided by the CMB group at the Helmholtz Zentrum München[4].
The Gillespie algorithm is implemented with the direct method in MATLAB.
The image fluorescence quantification was done by Michael Schwarzfischer and Adam Filipczyk with Schwarzfischer's tool **AMT** (Aided Manual Tracking), which normalizes the images for different background intensities at the margins and over time and calculates the intensity of a cell by subtracting the normalized background from the cell's pixel intensity. The normalization is in one position and takes into account both the unequal light distribu-

[2]http://www.mathworks.com
[3]http://www.yWorks.com
[4]http://www.helmholtz-muenchen.de/en/cmb/research/tools-databases/odefy

tion of the microscope lamp within one picture as well as the bleaching of the background over time.
Literature search was done with Google Scholar[5] and the PubMed database[6]. Text mining was performed with the Excerbt system [111] and the LitInspector tool by Genomatix [53].
The computer used for calculations and writing this thesis is a 32-bit notebook with 2Gb RAM and a 1.7GHz single core processor.

[5]http://scholar.google.com
[6]http://www.ncbi.nlm.nih.gov/pubmed

Bibliography

[1] Acar, M., Becskei, A., and van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435(7039):228–32, 2005. ISSN 1476-4687.

[2] Aiba, K., Nedorezov, T., Piao, Y., Nishiyama, A., Matoba, R., Sharova, L.V., Sharov, A.a., Yamanaka, S., Niwa, H., and Ko, M.S.H. Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. *DNA Research*, 16(1):73–80, 2009. ISSN 1756-1663.

[3] Albert, R. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *Journal of Theoretical Biology*, 223(1):1–18, 2003. ISSN 00225193.

[4] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–61, 2007. ISSN 1471-0056.

[5] Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development*, 17(1):126–40, 2003. ISSN 0890-9369.

[6] Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jø rgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*, 8(5):532–8, 2006. ISSN 1465-7392.

[7] Babaie, Y., Herwig, R., Greber, B., Brink, T.C., Wruck, W., Groth, D., Lehrach, H., Burdon, T., and Adjaye, J. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells*, 25(2):500–10, 2007. ISSN 1066-5099.

[8] Baharvand, H., Fathi, A., van Hoof, D., and Salekdeh, G.H. Concise review: trends in stem cell proteomics. *Stem cells*, 25(8):1888–903, 2007. ISSN 1066-5099.

[9] Becskei, A. and Serrano, L. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–3, 2000. ISSN 0028-0836.

[10] Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., and Lander, E.S. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–26, 2006. ISSN 0092-8674.

[11] Bharucha-Reid, A. *Elements of the Theory of Markov Processes and its Applications*. McGraw-Hill, New York, 1960.

[12] Bloechl, F., Theis, F.J., Vega-Redondo, F., and Fisher, E.O.N. Vertex Centralities in Input-Output Networks Reveal the Structure of Modern Economies. *Physical Review E*, pages 1–8, 2010.

[13] Boer, B., Kopp, J., Mallanna, S., Desler, M., Chakravarthy, H., Wilder, P.J., Bernadt, C., and Rizzino, A. Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Research*, 35(6):1773–86, 2007. ISSN 1362-4962.

[14] Bornholdt, S. Less Is More in Modeling Large Genetic Networks. *Science*, 310:449–51, 2005.

[15] Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.a., Jaenisch, R., and Young, R.a. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–56, 2005. ISSN 0092-8674.

[16] Boyer, L.A., Mathur, D., and Jaenisch, R. Molecular control of pluripotency. *Current Opinion in Genetics & Development*, 16(5):455–62, 2006. ISSN 0959-437X.

[17] Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell*, 2(2):151–9, 2008. ISSN 1875-9777.

[18] Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.a., and Vallier, L. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448(7150):191–5, 2007. ISSN 1476-4687.

[19] Brown, K., Legros, S., Artus, J., Doss, M.X., Khanin, R., Hadjantonakis, A.K., and Foley, A. A comparative analysis of extra-embryonic endoderm cell lines. *PloS ONE*, 5(8):e12016, 2010. ISSN 1932-6203.

[20] Bruce, A.W. and Zernicka-Goetz, M. Developmental control of the early mammalian embryo: competition among heterogeneous cells that biases cell fate. *Current Opinion in Genetics & Development*, 2010. ISSN 1879-0380.

[21] Cai, J., Chen, J., Liu, Y., Miura, T., Luo, Y., Loring, J.F., Freed, W.J., Rao, M.S., and Zeng, X. Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells*, 24(3):516–30, 2006. ISSN 1066-5099.

[22] Cai, K.Q., Capo-Chichi, C.D., Rula, M.E., Yang, D.H., and Xu, X.X. Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. *Developmental Dynamics*, 237(10):2820–9, 2008. ISSN 1058-8388.

[23] Campbell, P.A., Perez-Iratxeta, C., Andrade-Navarro, M.a., and Rudnicki, M.a. Oct4 targets regulatory nodes to modulate stem cell function. *PloS ONE*, 2(6):e553, 2007. ISSN 1932-6203.

[24] Carter, M.G., Sharov, A.A., VanBuren, V., Dudekula, D.B., Carmack, C.E., Nelson, C., and Ko, M.S.H. Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biology*, 6(7):R61, 2005. ISSN 1465-6914.

[25] Chambers, I. The Molecular Basis of Pluripotency in Mouse Embryonic Stem Cells. *Cloning and Stem Cells*, 6(4):386–391, 2004.

[26] Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–55, 2003. ISSN 0092-8674.

[27] Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173):1230–4, 2007. ISSN 1476-4687.

[28] Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental Cell*, 10(5):615–24, 2006. ISSN 1534-5807.

[29] Chen, L., Yabuuchi, A., Eminli, S., Takeuchi, A., Lu, C.W., Hochedlinger, K., and Daley, G.Q. Cross-regulation of the Nanog and Cdx2 promoters. *Cell Research*, 19(9):1052–61, 2009. ISSN 1748-7838.

[30] Chen, X., Fang, F., Liou, Y.C., and Ng, H.H. Zfp143 regulates Nanog through modulation of Oct4 binding. *Stem Cells*, 26(11):2759–67, 2008. ISSN 1549-4918.

[31] Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L., and Ng, H.H. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–17, 2008. ISSN 1097-4172.

[32] Chew, J.l., Loh, Y.h., Zhang, W., Chen, X., Tam, W.l., Yeap, L.s., Li, P., Ang, Y.s., Lim, B., and Robson, P. Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4 / Sox2 Complex in Embryonic Stem Cells. *Molecular and Cellular Biology*, 25(14):6031–6046, 2005.

[33] Chickarmane, V. and Peterson, C. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PloS ONE*, 3(10):e3478, 2008. ISSN 1932-6203.

[34] Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., Khvorostov, I., Ott, V., Grunstein, M., Lavon, N., Benvenisty, N., Croce, C.M., Clark, A.T., Baxter, T., Pyle, A.D., Teitell, M.a., Pelegrini, M., Plath, K., and Lowry, W.E. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell*, 5(1):111–23, 2009. ISSN 1875-9777.

[35] Ciaudo, C., Servant, N., Cognat, V., Sarazin, A., Kieffer, E., Viville, S., Colot, V., Barillot, E., Heard, E., and Voinnet, O. Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genetics*, 5(8):e1000620, 2009. ISSN 1553-7404.

[36] Cockburn, K. and Rossant, J. Review series Making the blastocyst : lessons from the mouse. *The Journal of Clinical Investigation*, 120(4):995–1003, 2010.

[37] Cookson, N.A., Cookson, S.W., Tsimring, L.S., and Hasty, J. Cell cycle-dependent variations in protein concentration. *Nucleic Acids Research*, 38(8):2676–2681, 2009. ISSN 1362-4962.

[38] Corish, P. and Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering*, 12(12):1035–40, 1999. ISSN 0269-2139.

[39] Daheron, L., Opitz, S.L., Zaehres, H., Lensch, W.M., Andrews, P.W., Itskovitz-Eldor, J., and Daley, Q. LIF/STAT3 Signaling Fails to Maintain Self-Renewal of Human Embryonic Stem Cells. *Stem Cells*, 22:770–778, 2004.

[40] Darr, H., Mayshar, Y., and Benvenisty, N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development*, 133(6):1193–201, 2006. ISSN 0950-1991.

[41] Davidich, M.I. and Bornholdt, S. Boolean network model predicts cell cycle sequence of fission yeast. *PloS ONE*, 3(2):e1672, 2008. ISSN 1932-6203.

[42] De Miguel, M.P., Fuentes-Julián, S., and Alcaina, Y. Pluripotent Stem Cells: Origin, Maintenance and Induction. *Stem Cell Reviews*, pages 633–649, 2010. ISSN 1558-6804.

[43] Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.F., Songyang, Z., Thomson, J.a., and Zwaka, T.P. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell*, 133(7):1162–74, 2008. ISSN 1097-4172.

[44] Eilken, H.M., Nishikawa, S.I., and Schroeder, T. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature*, 457(7231):896–900, 2009. ISSN 1476-4687.

[45] Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T., and Sasai, Y. Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature*, 472(7341):51–56, 2011. ISSN 0028-0836.

[46] Eldar, A. and Elowitz, M.B. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010. ISSN 0028-0836.

[47] Enver, T., Pera, M., Peterson, C., and Andrews, P.W. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell*, 4(5):387–97, 2009. ISSN 1875-9777.

[48] Evans, M.J. and Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156, 1981. ISSN 0028-0836.

[49] Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–31, 2006. ISSN 1367-4811.

[50] Fazzio, T.G. and Panning, B. Control of embryonic stem cell identity by nucleosome remodeling enzymes. *Current Opinion in Genetics & Development*, 20(5):500–4, 2010. ISSN 1879-0380.

[51] Ferri, A.L.M., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S., and Nicolis, S.K. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development*, 131(15):3805–19, 2004. ISSN 0950-1991.

[52] Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell*, 2(2):160–9, 2008. ISSN 1875-9777.

[53] Frisch, M., Klocke, B., Haltmeier, M., and Frech, K. LitInspector: literature and signal transduction pathway mining in PubMed abstracts. *Nucleic Acids Research*, 37(Web Server issue):W135–40, 2009. ISSN 1362-4962.

[54] Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J.i., and Niwa, H. Differentiation of embryonic stem cells is induced by GATA factors. *Genes & Development*, 16(7):784–9, 2002. ISSN 0890-9369.

[55] Gan, Q., Yoshida, T., McDonald, O.G., and Owens, G.K. Concise review: epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells. *Stem Cells*, 25(1):2–9, 2007. ISSN 1066-5099.

[56] Gangaraju, V.K. and Lin, H. MicroRNAs: key regulators of stem cells. *Nature Reviews Molecular Cell Biology*, 10(2):116–25, 2009. ISSN 1471-0080.

[57] Gillespie, D.T. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[58] Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J., and Rao, M.S. Differences between human and mouse embryonic stem cells. *Developmental Biology*, 269(2):360–80, 2004. ISSN 0012-1606.

[59] Giordano, A. and Galderisi, U. From the Laboratory Bench to the Patient ' s Bedside : An Update on Clinical Trials With Mesenchymal Stem Cells. *Journal of Cellular Physiology*, (10):27–35, 2007.

[60] Glauche, I., Herberg, M., and Roeder, I. Nanog variability and pluripotency regulation of embryonic stem cells–insights from a mathematical model analysis. *PloS ONE*, 5(6):e11238, 2010. ISSN 1932-6203.

[61] Graf, T. and Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell*, 3(5):480–3, 2008. ISSN 1875-9777.

[62] Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental Cell*, 18(4):675–85, 2010. ISSN 1878-1551.

[63] Hamdoun, A. and Epel, D. Embryo stability and vulnerability in an always changing world. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1745–1750, 2007.

[64] Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., and Schöler, H.R. Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. *Cell*, 1:1–11, 2010. ISSN 1097-4172.

[65] Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., Gao, Q., Kim, J., and Jaenisch, R. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell*, 4(6):513–24, 2009. ISSN 1875-9777.

[66] Hayashi, K., Chuva de Sousa Lopes, S.M., Kaneda, M., Tang, F., Hajkova, P., Lao, K., O'Carroll, D., Das, P.P., Tarakhovsky, A., Miska, E.a., and Surani, M.A. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PloS ONE*, 3(3):e1738, 2008. ISSN 1932-6203.

[67] Hinkelmann, F. and Laubenbacher, R. Boolean Models of Bistable Biological Systems. *Bioinformatics*, pages 1–17, 2009.

[68] Hochedlinger, K. and Plath, K. Epigenetic reprogramming and induced pluripotency. *Development*, 136(4):509–23, 2009. ISSN 0950-1991.

[69] Honda, H., Motosugi, N., Nagai, T., Tanemura, M., and Hiiragi, T. Computer simulation of emerging asymmetry in the mouse blastocyst. *Development*, 135(8):1407–14, 2008. ISSN 0950-1991.

[70] Houbaviy, H.B., Murray, M.F., and Sharp, P.a. Embryonic stem cell-specific MicroRNAs. *Developmental Cell*, 5(2):351–8, 2003. ISSN 1534-5807.

[71] Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–62, 2009. ISSN 1477-9129.

[72] Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Physical Review Letters*, 94(12):1–4, 2005. ISSN 0031-9007.

[73] Huh, D. and Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature genetics*, (December), 2010. ISSN 1546-1718.

[74] Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. Dissecting self-renewal in stem cells with RNA interference. *Nature*, 442(7102):533–8, 2006. ISSN 1476-4687.

[75] Ivey, K.N. and Srivastava, D. MicroRNAs as Regulators of Differentiation and Cell Fate Decisions. *Cell Stem Cell*, 7(1):36–41, 2010. ISSN 19345909.

[76] Jaenisch, R. and Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*, 132(4):567–82, 2008. ISSN 1097-4172.

[77] James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. TGF-beta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*, 132(6):1273–82, 2005. ISSN 0950-1991.

[78] Kalmar, T., Lim, C., Hayward, P., Muñoz Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biology*, 7(7):e1000149, 2009. ISSN 1545-7885.

[79] Kauffman, S. Homeostasis and Differentiation in Random Genetic Control Networks. *Nature*, 224:177–8, 1969.

[80] Kauffman, S. *Self orginazation and adaptation in complex systems*. Oxford University Press, New York, 1993.

[81] Kauffman, S. *At home in the universe*. Oxford University Press, New York, 1995.

[82] Kauffman, S. A proposal for using the ensemble approach to understand genetic regulatory networks. *Journal of Theoretical Biology*, 230(4):581–90, 2004. ISSN 0022-5193.

[83] Kaufmann, B.B., Yang, Q., Mettetal, J.T., and van Oudenaarden, A. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biology*, 5(9):e239, 2007. ISSN 1545-7885.

[84] Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes & Development*, 19(10):1129–55, 2005. ISSN 0890-9369.

[85] Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, 132(6):1049–61, 2008. ISSN 1097-4172.

[86] Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., Yabuuchi, A., Takeuchi, A., Cunniff, K.C., Hongguang, H., Mckinney-Freeman, S., Naveiras, O., Yoon, T.J., Irizarry, R.a., Jung, N., Seita, J., Hanna, J., Murakami, P., Jaenisch, R., Weissleder, R., Orkin, S.H., Weissman, I.L., Feinberg, a.P., and Daley, G.Q. Epigenetic memory in induced pluripotent stem cells. *Nature*, 467:285–92, 2010. ISSN 0028-0836.

[87] Kirkpatrick, S. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

[88] Kirouac, D.C., Madlambayan, G.J., Yu, M., Sykes, E.a., Ito, C., and Zandstra, P.W. Cell-cell interaction networks regulate blood stem and progenitor cell fate. *Molecular Systems Biology*, 5(293):293, 2009. ISSN 1744-4292.

[89] Klemm, K. and Bornholdt, S. Stable and unstable attractors in Boolean networks. *Physical Review E*, 72(5):1–4, 2005. ISSN 1539-3755.

[90] Komorowski, M., Finkenstädt, B., and Rand, D. Using a Single Fluorescent Reporter Gene to Infer Half-Life of Extrinsic Noise and Other Parameters of Gene Expression. *Biophysical journal*, 98(12):2759–2769, 2010. ISSN 1542-0086.

[91] Krawitz, P. and Shmulevich, I. Entropy of complex relevant components of Boolean networks. *Physical Review E*, 76(3):1–7, 2007. ISSN 1539-3755.

[92] Krumsiek, J., Marr, C., Schroeder, T., and Theis, F.J. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network, 2011.

[93] Kuijk, E.W., de Gier, J., Lopes, S.M.C.D.S., Chambers, I., van Pelt, A.M.M., Colenbrander, B., and Roelen, B.a.J. A distinct expression pattern in mammalian testes indicates a conserved role for NANOG in spermatogenesis. *PloS ONE*, 5(6):e10987, 2010. ISSN 1932-6203.

[94] Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*, 134(16):2895–902, 2007. ISSN 0950-1991.

[95] Lacham-Kaplan, O., Chy, H., and Trounson, A. Testicular cell conditioned medium supports differentiation of embryonic stem cells into ovarian structures containing oocytes. *Stem cells*, 24(2):266–73, 2006. ISSN 1066-5099.

[96] Lanner, F. and Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development*, 137(20):3351–60, 2010. ISSN 1477-9129.

[97] Lau, F., Ahfeldt, T., Osafune, K., Akustsu, H., and Cowan, C.a. Induced pluripotent stem (iPS) cells: an up-to-the-minute review. *F1000 biology reports*, 1(November):1–7, 2009. ISSN 1757-594X.

[98] Lensch, M.W., Daheron, L., and Schlaeger, T.M. Pluripotent stem cells and their niches. *Stem Cell Reviews*, 2(3):185–201, 2006. ISSN 1550-8943.

[99] Lestas, I., Vinnicombe, G., and Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178, 2010. ISSN 0028-0836.

[100] Lindner, B. Effects of noise in excitable systems. *Physics Reports*, 392(6):321–424, 2004. ISSN 03701573.

[101] Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K.Y., Sung, K.W., Lee, C.W.H., Zhao, X.D., Chiu, K.P., Lipovich, L., Kuznetsov, V.a., Robson, P., Stanton, L.W., Wei, C.L., Ruan, Y., Lim, B., and Ng, H.H. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics*, 38(4):431–40, 2006. ISSN 1061-4036.

[102] Longtin, A. Stochastic resonance in neuron models: endogenous stimulation revisited. *Journal of Statistical Physics*, 70(1):309–327, 1993. ISSN 1539-3755.

[103] Lutter, D. A Boolean model for the differentiation of endoderm, mesoderm and ectoderm, 2011.

[104] MacArthur, B.D., Please, C.P., and Oreffo, R.O.C. Stochasticity and the molecular mechanisms of induced pluripotency. *PloS ONE*, 3(8):e3086, 2008. ISSN 1932-6203.

[105] Maheshri, N. and O'Shea, E.K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annual Review of Biophysics and Biomolecular Structure*, 36:413–34, 2007. ISSN 1056-8700.

[106] Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., Calabrese, J.M., Dennis,

L.M., Volkert, T.L., Gupta, S., Love, J., Hannett, N., Sharp, P.a., Bartel, D.P., Jaenisch, R., and Young, R.a. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, 134(3):521–33, 2008. ISSN 1097-4172.

[107] Martin, G.R. Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells. *Proceedings of the National Academy of Sciences*, 78(12):7634–7638, 1981. ISSN 0027-8424.

[108] Martinez, N.J. and Gregory, R.I. MicroRNA Gene Regulatory Pathways in the Establishment and Maintenance of ESC Identity. *Cell Stem Cell*, 7(1):31–35, 2010. ISSN 19345909.

[109] Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.a., and Ko, M.S.H. Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PloS ONE*, 1(1):e26, 2006. ISSN 1932-6203.

[110] Mattout, A. and Meshorer, E. Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Current Opinion in Cell Biology*, 22(3):334–341, 2010. ISSN 1879-0410.

[111] Mewes, H.W., Ruepp, A., Theis, F., Rattei, T., Walter, M., Frishman, D., Suhre, K., Spannagl, M., Mayer, K.F.X., Stümpflen, V., and Antonov, A. MIPS: curated databases and comprehensive secondary data resources in 2010. *Nucleic Acids Research*, 39:220–224, 2010. ISSN 1362-4962.

[112] Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. Dissecting direct reprogramming through integrative genomic analysis. *Nature*, 454(7200):49–55, 2008. ISSN 1476-4687.

[113] Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., and Bernstein, B.E. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153):553–60, 2007. ISSN 1476-4687.

[114] Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113(5):631–42, 2003. ISSN 0092-8674.

[115] Morris, S.A., Teo, R.T.Y., Li, H., Robson, P., Glover, D.M., and Zernicka-Goetz, M. Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14):6364–9, 2010. ISSN 1091-6490.

[116] Muers, M. Stem cells: The highs and lows of pluripotency. *Nature Reviews Genetics*, 10(9):590–591, 2009. ISSN 1471-0056.

[117] Mullen, E.M., Gu, P., and Cooney, A.J. Nuclear Receptors in Regulation of Mouse ES Cell Pluripotency and Differentiation. *PPAR Research*, 2007:61563, 2007. ISSN 1687-4757.

[118] Müller, F.J., Laurent, L.C., Kostka, D., Ulitsky, I., Williams, R., Lu, C., Park, I.H., Rao, M.S., Shamir, R., Schwartz, P.H., Schmidt, N.O., and Loring, J.F. Regulatory networks define phenotypic classes of human stem cell lines. *Nature*, 455(7211):401–5, 2008. ISSN 1476-4687.

[119] Mullin, N.P., Yates, A., Rowe, A.J., Nijmeijer, B., Colby, D., Barlow, P.N., Walkinshaw, M.D., and Chambers, I. The pluripotency rheostat Nanog functions as a dimer. *The Biochemical Journal*, 411(2):227–31, 2008. ISSN 1470-8728.

[120] Munsky, B., Trinh, B., and Khammash, M. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(318):318, 2009. ISSN 1744-4292.

[121] Narayanan, S. *The Betweenness Centrality Of Biological Networks*. Ph.D. thesis, University of Virginia, Blacksburg, 2005.

[122] Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3):379–91, 1998. ISSN 0092-8674.

[123] Nichols, J. and Smith, A. Naive and primed pluripotent states. *Cell Stem Cell*, 4(6):487–92, 2009. ISSN 1875-9777.

[124] Niepel, M., Spencer, S.L., and Sorger, P.K. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Current Opinion in Chemical Biology*, 13(5-6):556–61, 2009. ISSN 1879-0402.

[125] Niwa, H., Miyazaki, J., and Smith, a.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(4):372–6, 2000. ISSN 1061-4036.

[126] Niwa, H. How is pluripotency determined and maintained? *Development (Cambridge, England)*, 134(4):635–46, 2007. ISSN 0950-1991.

[127] Niwa, H. Open conformation chromatin and pluripotency. *Genes & Development*, 21(21):2671–6, 2007. ISSN 0890-9369.

[128] Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*, 123(5):917–29, 2005. ISSN 0092-8674.

[129] Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. A signaling principle for the specification of the germ cell lineage in mice. *Cell*, 137(3):571–84, 2009. ISSN 1097-4172.

[130] Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31(1):69–73, 2002. ISSN 1061-4036.

[131] Pan, G., Li, J., Zhou, Y., Zheng, H., and Pei, D. A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *The FASEB journal*, 20(10):1730–2, 2006. ISSN 1530-6860.

[132] Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell*, 6(4):382–95, 2010. ISSN 1875-9777.

[133] Pasquinelli, A.E., Hunter, S., and Bracht, J. MicroRNAs: a developing story. *Current Opinion in Genetics & Development*, 15(2):200–5, 2005. ISSN 0959-437X.

[134] Paulsson, J. Models of stochastic gene expression. *Physics of Life Reviews*, 2(2):157–175, 2005. ISSN 15710645.

[135] Paulsson, J. Summing up the noise in gene networks. *Nature*, 427(6973):415–8, 2004. ISSN 1476-4687.

[136] Peterson, J.L. Petri Nets. *ACM*, 9(3), 1977. ISSN 1932-6203.

[137] Pfister, S., Steiner, K.a., and Tam, P.P.L. Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. *Gene Expression Patterns*, 7(5):558–73, 2007. ISSN 1567-133X.

[138] Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, 4(10):e309, 2006. ISSN 1545-7885.

[139] Raj, A. and van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26, 2008. ISSN 1097-4172.

[140] Rao, M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Developmental Biology*, 275(2):269–86, 2004. ISSN 0012-1606.

[141] Raser, J.M. and O'Shea, E.K. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–3, 2005. ISSN 1095-9203.

[142] Rieger, M.a., Hoppe, P.S., Smejkal, B.M., Eitelhuber, A.C., and Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science*, 325(5937):217–8, 2009. ISSN 1095-9203.

[143] Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.a., Vultur, A., Basu, D., Gimotty, P., Vogt, T., and Herlyn, M. A Temporarily Distinct Subpopulation of Slow-Cycling Melanoma Cells Is Required for Continuous Tumor Growth. *Cell*, 141(4):583–594, 2010. ISSN 00928674.

[144] Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. Gene regulation at the single-cell level. *Science*, 307(5717):1962–5, 2005. ISSN 1095-9203.

[145] Saez-Rodriguez, J., Simeoni, L., Lindquist, J.a., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U.U., Weismantel, R., Gilles, E.D., Klamt, S., and Schraven, B. A logical model provides insights into T cell receptor signaling. *PLoS Computational Biology*, 3(8):e163, 2007. ISSN 1553-7358.

[146] Saitou, M. Germ cell specification in mice. *Current Opinion in Genetics & Development*, 19(4):386–95, 2009. ISSN 1879-0380.

[147] Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine*, 10(1):55–63, 2004. ISSN 1078-8956.

[148] Schroeder, T. Imaging stem-cell-driven regeneration in mammals. *Nature*, 453(7193):345–51, 2008. ISSN 1476-4687.

[149] Sensebé, L., Krampera, M., Schrezenmeier, H., Bourin, P., and Giordano, R. Mesenchymal stem cells for clinical application. *Vox sanguinis*, 98(2):93–107, 2010. ISSN 1423-0410.

[150] Shahrezaei, V. and Swain, P.S. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45):17256–61, 2008. ISSN 1091-6490.

[151] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–504, 2003. ISSN 1088-9051.

[152] Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.a., Wong, K.K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., and Settleman, J. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80, 2010. ISSN 1097-4172.

[153] Sharova, L.V., Sharov, A.a., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research*, 16(1):45–58, 2009. ISSN 1756-1663.

[154] Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. Variability and memory of protein levels in human cells. *Nature*, 444(7119):643–6, 2006. ISSN 1476-4687.

[155] Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biology*, 6(10):e253, 2008. ISSN 1545-7885.

[156] Silva, J. and Smith, A. Capturing pluripotency. *Cell*, 132(4):532–6, 2008. ISSN 1097-4172.

[157] Singh, A.M., Hamazaki, T., Hankowski, K.E., and Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells*, 25(10):2534–42, 2007. ISSN 1549-4918.

[158] Smith, a.G. Embryo-derived stem cells: of mice and men. *Annual review of cell and developmental biology*, 17:435–62, 2001. ISSN 1081-0706.

[159] Smith, A. The battlefield of pluripotency. *Cell*, 123(5):757–60, 2005. ISSN 0092-8674.

[160] Smith, A. Design principles of pluripotency. *EMBO Molecular Medicine*, 1(5):251–4, 2009. ISSN 1757-4684.

[161] Smith, A., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336:688–90, 1988.

[162] Som, A., Harder, C., Greber, B., Siatkowski, M., Paudel, Y., Warsow, G., Cap, C., Schöler, H., and Fuellen, G. The PluriNetWork: An Electronic Representation of the Network Underlying Pluripotency in Mouse, and Its Applications. *PloS ONE*, 5(12):e15165, 2010. ISSN 1932-6203.

[163] Spiller, C.M., Wilhelm, D., and Koopman, P. Retinoblastoma 1 protein modulates XY germ cell entry into G1/G0 arrest during fetal development in mice. *Biology of Reproduction*, 82(2):433–43, 2010. ISSN 1529-7268.

[164] Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–50, 2006. ISSN 1476-4687.

[165] Swain, P.S., Elowitz, M.B., and Siggia, E.D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12795–800, 2002. ISSN 0027-8424.

[166] Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5):861–72, 2007. ISSN 0092-8674.

[167] Takahashi, K. and Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, 2006. ISSN 0092-8674.

[168] Tam, P.P. and Behringer, R.R. Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development*, 68(1-2):3–25, 1997. ISSN 0925-4773.

[169] Tang, F., Hajkova, P., Barton, S.C., Lao, K., and Surani, M.A. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Research*, 34(2):e9, 2006. ISSN 1362-4962.

[170] Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, 455(7216):1124–8, 2008. ISSN 1476-4687.

[171] Tesar, P.J., Chenoweth, J.G., Brook, F.a., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150):196–9, 2007. ISSN 1476-4687.

[172] Thattai, M. and van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8614–9, 2001. ISSN 0027-8424.

[173] Thomson, J.a. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*, 282(5391):1145–1147, 1998.

[174] Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*, 135(5):909–18, 2008. ISSN 0950-1991.

[175] Tyson, J. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003. ISSN 09550674.

[176] Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.H., Martinez, A., Rugg-Gunn, P., Brons, G., and Pedersen, R.a. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development*, 136(8):1339–49, 2009. ISSN 0950-1991.

[177] van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.a. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*, 6(4):369–81, 2010. ISSN 1875-9777.

[178] Villani, M., Barbieri, A., and Serra, R. A Dynamical Model of Genetic Networks for Cell Differentiation. *PLoS ONE*, 6(3):e17703, 2011. ISSN 1932-6203.

[179] Waddington, C.H. *The strategy of the genes*. Allen & Unwin, 1957.

[180] Wang, J., Levasseur, D.N., and Orkin, S.H. Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, 105(17):6326–31, 2008. ISSN 1091-6490.

[181] Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. A protein interaction network for pluripotency of embryonic stem cells. *Nature*, 444(7117):364–8, 2006. ISSN 1476-4687.

[182] Warsow, G., Greber, B., Falk, S.S.I., Harder, C., Siatkowski, M., Schordan, S., Som, A., Endlich, N., Scholer, H., Repsilber, D., Endlich, K., and Fuellen, G. ExprEssence - Revealing the essence of differential experimental data in the context of an interaction/regulation network. *BMC Systems Biology*, 4(1):164, 2010. ISSN 1752-0509.

[183] Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., and Lander, E.S. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–62, 2002. ISSN 0028-0836.

[184] Weissman, I.L. and Shizuru, J.a. The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood*, 112(9):3543–53, 2008. ISSN 1528-0020.

[185] Wittmann, D.M., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D.a., Klamt, S., and Theis, F.J. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Systems Biology*, 3:98, 2009. ISSN 1752-0509.

[186] Wittmann, D.M., Marr, C., and Theis, F.J. Biologically meaningful update rules increase the critical connectivity of generalized Kauffman networks. *Journal of theoretical biology*, 266(3):436–448, 2010. ISSN 1095-8541.

[187] Xu, H., Schaniel, C., Lemischka, I.R., and Ma'ayan, A. Toward a complete in silico, multi-layered embryonic stem cell regulatory network. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 2(6):708–33, 2010. ISSN 1939-005X.

[188] Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.a., and Kosik, K.S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*, 137(4):647–58, 2009. ISSN 1097-4172.

[189] Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature*, 460(7251):49–52, 2009. ISSN 1476-4687.

[190] Yamanaka, Y., Lanner, F., and Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*, 137(5):715–24, 2010. ISSN 1477-9129.

[191] Yasunaga, M., Tada, S., Torikai-Nishikawa, S., Nakano, Y., Okada, M., Jakt, L.M., Nishikawa, S., Chiba, T., Era, T., and Nishikawa, S.I. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nature Biotechnology*, 23(12):1542–50, 2005. ISSN 1087-0156.

[192] Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–23, 2008. ISSN 1476-4687.

[193] Yu, J. and Thomson, J.A. Pluripotent stem cell lines. *Genes & Development*, 22(15):1987–97, 2008. ISSN 0890-9369.

[194] Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y.I., Aburatani, H., Ura, K., Koseki, H., Niwa, H., and Nishinakamura, R. Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression. *Stem cells*, 27(4):796–805, 2009. ISSN 1549-4918.

[195] Zheng, Z., de Iongh, R.U., Rathjen, P.D., and Rathjen, J. A requirement for FGF signalling in the formation of primitive streak-like intermediates from primitive ectoderm in culture. *PloS ONE*, 5(9):e12555, 2010. ISSN 1932-6203.

[196] Zipori, D. The nature of stem cells: state rather than entity. *Nature Reviews Genetics*, 5(11):873–8, 2004. ISSN 1471-0056.