HUMBOLDT-UNIVERSITÄT ZU BERLIN

MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT I

INSTITUT FÜR BIOLOGIE

# Diplomarbeit

ZUM ERWERB DES AKADEMISCHEN GRADES

DIPLOM-BIOPHYSIKER

# *Mathematical Modeling of Stem Cell Reprogramming*

vorgelegt von

Till Scharp

geb. am 01.08.1985 in Berlin

angefertigt in der Arbeitsgruppe Theoretische Biophysik

am Institut für Biologie

Berlin, im Februar 2010

> “Wenn sie nun imstande wären, miteinander das Erblickte an- und durchzusprechen, glaubst du nicht, sie würden das, was sie da sehen, für das Seiende halten?”

Platon, Politeia VII, 515a

# Abstract

The following thesis treats of methods and procedures for the mathematical modeling of gene regulatory networks (GRNs) involved in the determination of pluripotency and self-renewal. It aims at mathematically explaining the processes of embryonic stem cell (ESC) differentiation and induced pluripotent stem cell (iPSC) reprogramming. We are presenting different versions of gene regulatory networks that rely on literature research, ChIP-on-chip and microarray data. These networks involve the core regulatory genes of pluripotency, epigenetic remodeling factors, microRNAs and genes from signaling pathways. The analysis of these networks is done by different mathematical means, the Boolean modeling method, the standardized qualitative dynamical approach and with systems of ordinary differential equations (ODEs) to give an insight into the dynamical behavior of the networks.

We analyzed 5 different network structures, the Chickarmane network, the Boyer core network, the epigenetic network model, the miRNA network model and the signaling pathway network model. For the first, we were able to reproduce some of the dynamics mentioned in the publication of Chickarmane and Peterson in 2008 but at the same time we showed the limitations of Boolean network models when it comes to more complex than switch like dynamics, e.g. biphasic behavior. With the small epigenetic model we were able to explain the importance of the appropriate timing of epigenetic reprogramming. The miRNA network model introduced a repressive switch applied to the core network of pluripotency that could provide possibilities why NANOG alone is not sufficient in reprogramming. Lastly, the pathway model could give an impression of the interplay of the branches of TGFß and FGF2 in reprogramming and differentiation.

Overall, by the mathematical analysis of the different mechanistic networks we were able to reconstruct known dynamics of reprogramming and found possible reasons for the low reprogramming efficiencies and long reprogramming timing. It will be of interest one day to merge these models in a systems biology approach to get a holistic view of pluripotency and iPSC reprogramming.

# Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der mathematischen Modellierung von genregulatorischen Netzwerken, die bei der Bestimmung von Pluripotenz und Selbsterneuerung eine Rolle spielen. Ziel der Arbeit ist die mathematische Analyse der Prozesse, die bei der Differenzierung von embryonalen Stammzellen (embryonic stem cells, ESCs) und der Reprogrammierung von differenzierten zu induzierten pluripotenten Stammzellen (induced pluripotent stem cells, iPSCs) stattfinden. Wir stellen verschiedene genregulatorische Netzwerke vor, die auf Literaturrecherchen, ChIP-on-chip- und Microarray-Daten beruhen. Diese beinhalten kernregulatorische Gene der Pluripotenz, epigenetische Faktoren, microRNAs und Gene aus Signalwegen. Die Analyse dieser Netzwerke und ihres dynamischen Verhaltens erfolgt mittels mathematischer Methoden, der Booleschen Modellierung, der standardisierten qualitativen dynamischen Systeme und mittels Systemen gewöhnlicher Differentialgleichungen.

Wir analysierten 5 verschiedene Netzwerkstrukturen: das Chickarmane Netzwerk, das Boyer Kernnetzwerk, unser epigenetisches Modell, ein mit miRNAs erweitertes Netzwerk und ein System, das bestimmte Signalwege beinhaltet. Beim ersten waren wir in der Lage die Dynamiken von Chickarmane und Peterson 2008 teilweise zu reproduzieren, stießen jedoch auf die Einschränkungen der Booleschen Modellierung bei komplizierteren Dynamiken, z.B. biphasischem Verhalten. Mittels unseres epigenetischen Netzwerks konnten wir die Wichtigkeit der zeitlichen Abfolge der epigenetischen Reprogrammierung nachweisen. Mittels der miRNAs führten wir einen repressiven Schaltermechanismus ein, der auf das Kernnetzwerk ausgeübt wird. Dieser konnte Möglichkeiten aufzeigen, weshalb NANOG alleine nicht ausreichend für die Reprogammierung ist. Zuletzt konnten wir mit dem Signalwegmodell einen Überblick über das Zusammenspiel der Zweige des TGFß und des FGF2 Weges bei der Reprogrammierung und der Differenzierung geben.

Wir konnten dank der mathematischen Analyse der verschiedenen Netzwerke bekannte Dynamiken des Prozesses rekonstruieren und Möglichkeiten für die geringe Effizienz der Reprogrammierung aufzeigen. Es wird außerdem in Zukunft von großem Interesse sein, die verschiedenen Modelle in einer systembiologisch ganzheitlichen Sicht der Pluripotenz und iPSC Reprogrammierung zusammenzuführen.

# Acknowledgments

I would like to thank Prof. Edda Klipp for her support and supervision and for making this work possible. Furthermore I'd like to thank Dr. James Adjaye, Dr. Alexander Hahn, Dr. Ralf Mrowka, Dr. Miguel Andrade and their groups for providing me with data and helpful information. Moreover I thank all the group of theoretical biophysics from Prof. Edda Klipp for being so nice and always there to help me with every kind of arising questions. I'd especially like to thank Max Flöttmann from the group for helping me develop my ideas and constantly supporting me in all my problems. I'd really like to thank all my friends who were always there for me during the good and the bad times I had writing this thesis. Special thanks go to my roommates Daniel Koschyk and Sven Schotte for supporting me and saving me time with everything that had to do with the apartment. Last but not least the biggest thanks goes out to my parents for always supporting me in the pursuit of my work and all aspects of my life. Thank you all very much!

# Contents

# Abbreviations

3’ UTR = 3’ untranslated region

ALS = amyotropic lateral sclerosis

bFGF = basic fibroblast growth factor

BMP = bone morphogenetic protein

bp(s) = base pair(s)

CGBP = CpG binding protein

DNA = Desoxyribonucleic acid

ECCs = embryonic carcinoma cells

GRNs = gene regulatory networks

HDAC = histone deacetylase

(h)/(m)ESCs = (human)/(mouse) embryonic stem cells

ICM = inner cell mass

Id = inhibitor of differentiation

iPSCs = induced pluripotent stem cells

JAK = Janus-associated tyrosine kinase

LIF = leukemia inhibitory factor

MEFs = mouse embryonic fibroblasts

mRNA = messenger ribonucleic acid

miRNA = micro ribonucleic acid

ncRNA = non-coding ribonucleic acid

p.c. = post conceptionem

PPEs = promoter proximal element

RA = retinoic acid

REST = RE1-silencing transcription factor

rRNA = ribosomal ribonucleic acid

siRNA = small interfering ribonucleic acid

snRNA = small nuclear ribonucleic acid

snoRNA = small nucleolar ribonucleic acid

STAT = signal transducers and activators of transcription

ODEs = ordinary differential equations

TF = transcription factor

# 1. Introduction

## 1.1. Presentation of the subject

The investigation and characterization of embryonic stem cells (ESCs) is one of the most promising and most relevant fields of today's research. The main focus is therein put on the development of medical applications for the treatment of many degenerative diseases, i.e. diseases in which the affected tissues will progressively degenerate with time. The reason for their attractivity is that ESCs have the potential to develop into any cell of the organism, the so called pluripotency, when the right triggers are applied and to self-renew indefinitely in vivo and in vitro under the appropriate conditions.

As the direct acquisition of human embryonic stem cells (hESCs) requires the killing of a human embryo which is ethically controversial, the recently developed concept of nuclear reprogramming first performed by Takahashi and Yamanaka in 2006 [1], is becoming more and more attractive for medical therapy. In their work, Yamanaka and Takahashi introduced genes coding for the four transcription factors Oct4, Sox2, Klf4 and C-myc via viral transfection into mouse embryonic fibroblasts (MEFs). Thereby, they induced the reprogramming of these terminally differentiated cells into cells that strongly resembled ESCs from a morphological and genetic point of view and which they called induced pluripotent stem cells (iPSCs).

As degenerative diseases have mainly genetic issues, the ideal workflow of a hypothetical therapeutic approach would be to extract terminally differentiated cells from a patient with the degenerative disease, e.g. skin cells [2] or cells from adipose tissue [3], to reprogram them in vitro into iPSCs and to repair the genetic defect in this culture. As a next step, the repaired iPSCs could be differentiated into the cell lineage from the affected tissue and retransplanted into the patient (see Figure 1.1). This strategy offers large advantages compared to the "conventional" stem cell therapy where unspecific ESCs from an existing cell line would be used. In first place, because the transplanted cells originate from a patient's graft, immune rejection could be prevented. Moreover, as the iPSCs are patient specific, diseases could be evaluated more individually and not only for model ESCs which may not be able to reflect the vast majority of patient's genetic defects.

## 1.2. State of the art in iPSC research: possibilities and problems

The reprogramming method of Yamanaka in 2006 was a big breakthrough in the field of ESC therapy. Nevertheless, the arising criticism about the use of the transcription factor c-myc which is commonly known to be an oncogene [5], the low reprogramming efficiency and the involvement of viruses randomly integrating and remaining in the genome after transfection, even though they are silenced, led to the development of further reprogramming techniques.

In 2007, Yu et al. used a slightly different cocktail consisting of *Sox2*, *Nanog*, *Lin28* and *Pou5f1*, the ladder being the name of the gene coding for Oct4, to reprogram differentiated cells into iPSCs thus avoiding the *C-myc* oncogene although the efficiency was further decreased compared to the Yamanaka cocktail [6]. However, an enhancement of the efficiency was attained by Liao et al. in 2008 when they combined the two cocktails to a six factor comprising combination of *Pou5f1*, *Sox2*, *Nanog*, *Klf4*, *C-myc* and *Lin28* [7]. Another very effective method to enhance the efficiency is the use of small molecules in addition to the reprogramming factors. Valproic acid, a histone deacetylase (HDAC) inhibitor is one of these small molecules that is able to increase the efficiency by a factor of approximately 100 [8]. A virus-free induction method was performed by Okita et al in 2008 and Kaji et al in 2009, where they used a plasmid for the transfection with the reprogramming factors that could be excised once the reprogramming was done [9, 10]. In another approach some reprogramming factors were replaced by small molecule compounds but viruses for the transfection with the remaining factors were still necessary [11, 12]. In 2009, Kim et al showed that human fibroblasts could be reprogrammed by the direct delivery of the reprogramming proteins fused to a cell penetrating peptide thereby avoiding the use of viruses or possibly harmful chemical compounds [13].

Another question that arose early in iPSC research is on the concern of the rather long reprogramming times. In fact, the reprogramming experiment proposed by Yamanaka and Takahashi and also the one of Yu et al. which were the two reprogramming methods adopted by most groups in this field needed at least 12 days to show the first iPS colonies [6] while full reprogramming took even longer up to 28 days or more. Ongoing theories to clarify this problems name epigenetic modifications or the low transfection efficiency as main reasons for this reprogramming timing. Very recent results published in December 2009 by Hanna et al. showed that assumingly it is not the absolute reprogramming time but the number of cell divisions taking place during the reprogramming that is responsible for the time of the process [14]. The time period can thus be associated with the epigenetic remodeling of the cell's DNA since this is one of the few processes that can be altered by the number of cell divisions. Simultaneously to the

remodeling of the endogenous core transcriptional factors of pluripotency that will be activated during the reprogramming, the exogenous retroviral genes are silenced epigenetically. Possible mechanisms for both processes will be discussed in 2.2.2 and later in the result section in 4.2 where we combined the concepts of epigenetic reprogramming and silencing of retroviral genes in order to give possible reasons for the long reprogramming times.

## 1.3. The biological origin of ESCs

To understand the origin and the purpose of ESCs, we will have a very simplified look at the embryogenesis of mammals especially humans, i.e. the process of development of an embryo after fertilization of an ovule (egg cell) by a spermatozoon.

Directly after the fertilization the egg cell (or zygote) undergoes rapid cell divisions without significant growth resulting in a cluster of cells called the morula. Approximately 4 days after fertilization, the blastocoel appears, a cavity filled with fluid and the morula becomes a blastula consisting of around 128 cells. When this blastula develops into the blastocycst, it is composed of the trophoblast forming the outer layer and of the inner cell mass (ICM) (see Figure 1.2) While the trophoblast represents extraembryonic tissue giving rise to the placenta, the amniotic sac and nutrients for the embryo, the ICM (or embryoblast) consisting of ESCs will develop into all the tissues of the later organism. At the stage of the ICM, the cells are pluripotent but not omni- or totipotent anymore because they lost their ability to give rise to extraembryonic tissue.

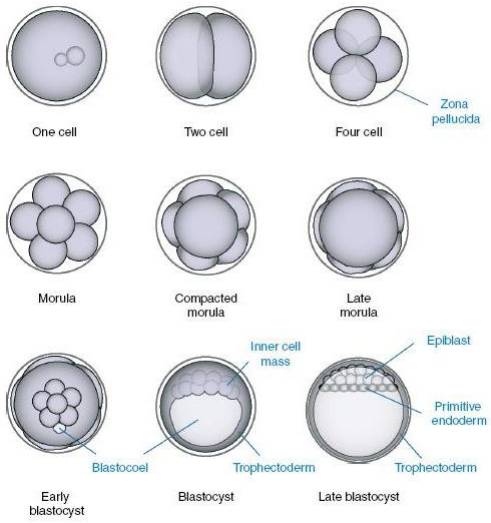


Figure 1.2.: First part of the mouse embryogenesis from the zygote until the late blastocyst stage [15]

At this stage of development, the blastocyst begins the implantation into the uterus via a part of the trophoblast while the embryoblast divides into the hypoblast which will form the extraembryonic endoderm giving rise to the yolk sac and into the epiblast. The epiblast then divides again into the embryonic epiblast and other epiblast cells forming the amniotic cavity. The embryonic epiblast will continue its development and perform the gastrulation approximately at day 16 after conception (post conceptionem p.c.). During that process cells of the embryonic epiblast migrate through the primitive streak in direction of the hypoblast replacing its cells progressively thus forming a new layer, the endoderm. While the original epiblast now forms the ectoderm layer, cells remaining between it and the newly created endoderm will form the mesoderm (see Figure 1.3). These 3 germ layers evolving from the embryonic epiblast will later develop into the internal organs of the organism during the organogenesis which takes place within the 3rd to 8th week of p.c.

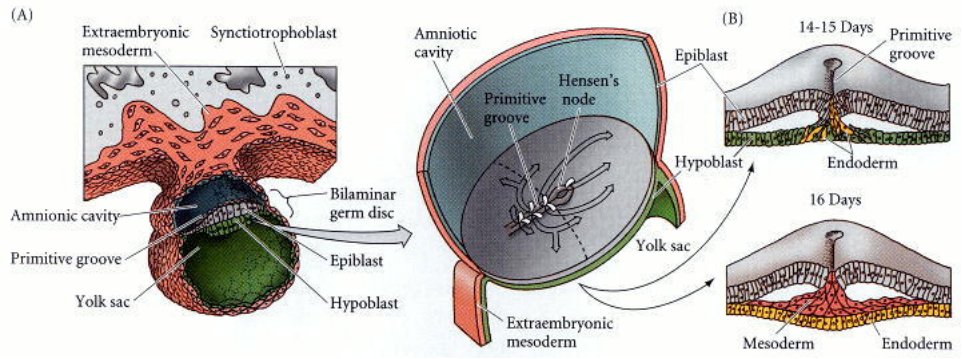


Figure 1.3.: The process of gastrulation giving rise to the three germ layers which will develop into all tissues of the later organism [developmental biology, Scott F Gilbert]

## 1.4. Aim of the work

The field of ESCs and iPSCs reprogramming is very young compared to others and thus scientific facts, although strongly emerging in the last years, are still scarce and a lot of unproved theories and unsolved problems surround it. The following work will give an overview of the possibilities and limits of mathematical modeling applied to several possible versions of gene regulatory networks (GRNs), that are likely to be involved in the reprogramming of differentiated cells into iPSCs, as a means to understand how the reprogramming factors act in multiple ways on the cell.

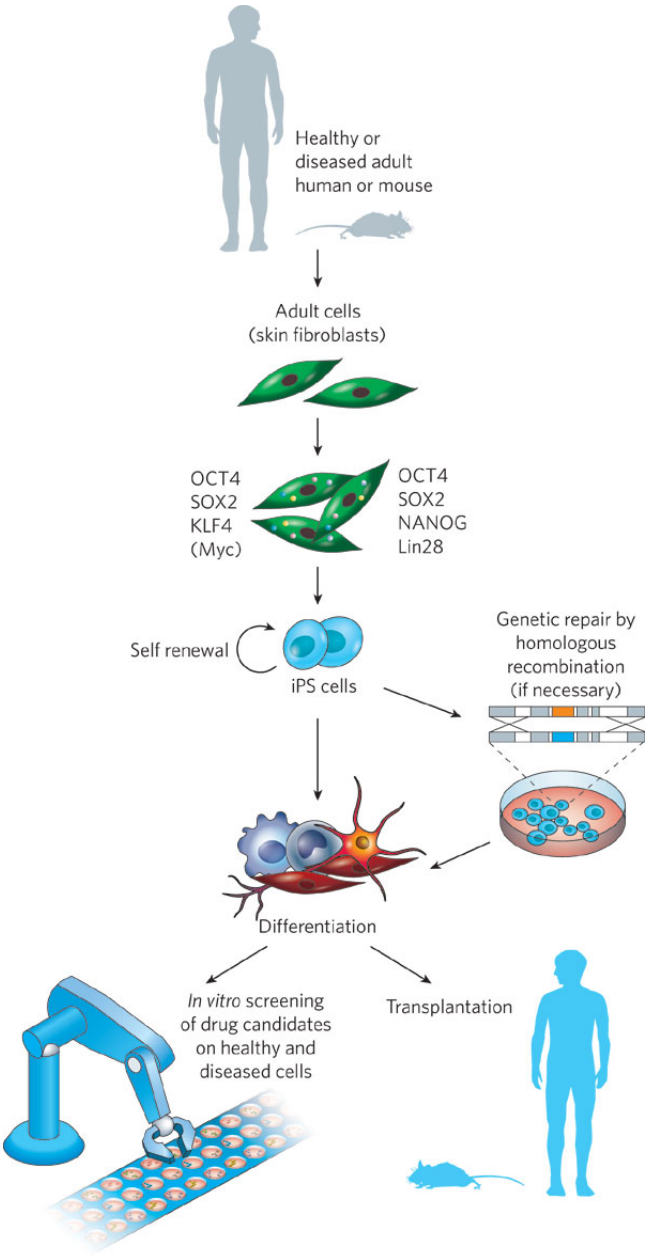


Figure 1.1.: Possible iPSC therapy [4]. Adult differentiated cells (e.g. fibroblasts) are taken from the diseased mouse or human, cultured in a dish and transfected with the reprogramming cocktail. The resulting iPSCs are then genetically manipulated to become healthy again, differentiated back to cells of the tissue in question and retransplanted into the patient, regenerating the unhealthy tissue.

# 2. Theoretical background

## 2.1. How to control protein expression on the genetic level

### 2.1.1. From genes to proteins: the central dogma of molecular biology

The genome encodes some ten thousands of genes whose gene products make cell survival and cellular functions possible. Since its first use in 1909 by Wilhelm Johannsen in *Limitations of natural selection on pure lines* [16], the definition of the word "gene" has constantly evolved with the findings and progress in the research. However until today, there are ongoing controversial discussions and there is no commonly accepted truth on this matter [17]. With a contemporary molecular definition by Lodish a gene corresponds to "the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide" [18]. By that definition, genes do not only code for mRNAs which can directly be translated into proteins but also for other types of RNA which are necessary to convert the mRNA into a functional protein e.g. rRNAs, snRNAs, snoRNAs or miRNAs which will be mentioned later on in this work again.

The concept of the central dogma of molecular cell biology was first introduced by Crick in 1958, elaborated in 1970 [19] and has evolved ever since. It mainly states that the flow of information between the species carrying information, i.e. DNA, RNA and proteins takes place in the direction of the protein but not vice versa [20] as shown in Figure 2.1. In the following we will give a short overview of the processes involved in this information flow to get an insight into protein biosynthesis. However we will not give the detailed molecular mechanisms in here, since they will be mentioned further in the next sections.

The DNA is localized in the nucleus of a cell where it is associated with proteins, mainly histones, to form the chromatin. This chromatin can take its transcriptionally active euchromatin or inactive heterochromatin conformation depending on the modifications of the DNA and the histones. When in an active conformation, it is accessible for the RNA polymerase II which can bind to the DNA and initiate its transcription into the corresponding pre-mRNA. The binding takes place at the promoter of the gene, i.e. an evolutionary conserved consensus sequence,

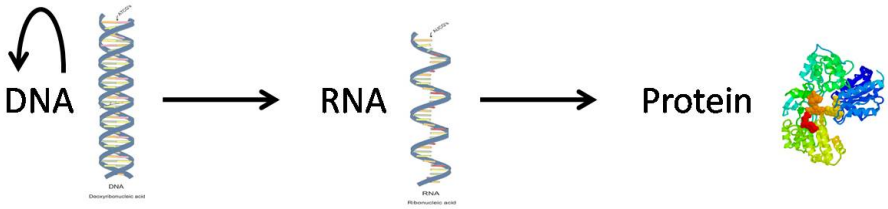


Figure 2.1.: The central dogma of molecular cell biology as envisioned by Watson in 1965 (inspired from [21])

such as the TATA box in Eukaryotes that lies 25-35 base pairs (bps) upstream of the transcription starting point. However, not all euchromatin is transcriptionally active which is due to the fact that the process of transcription depends furthermore on a large set of transcription factors (TFs) [18, 22]. There are two groups of transcription factors: general TFs are essential for building up the transcription initiation complex required for the transcription in general. The second group consists of specific TFs that can positively or negatively regulate the transcription of certain genes by mechanisms depicted in 1.4.2. In that section we will also discuss the role of transcription regulators that are not defined as TFs.

Following the transcription, the pre-mRNA is processed to become a mature mRNA and transported from the nucleus through the pores of the nuclear membrane into the cytoplasm. There, it can be translated by the ribosomes into polypeptides which can then fold spontaneously or mature in the endoplasmic reticulum and the Golgi apparatus to become the active protein. We will not mention the detailed mechanisms of mRNA processing, translation and protein folding, because it is not important for the subject of this thesis.

### 2.1.2. Mechanisms of eukaryotic gene control

In multicellular organisms, the most important aim of gene control is the execution of exact developmental decisions. The right gene must be regulated in the right cell at the right time during the development of numerous different cell types. Although gene control can occur on many different levels, most genes in higher Eukaryotes are regulated on the stage of transcription initiation [18].

#### Regulation of transcription initiation

As mentioned in 2.1.1, there are specific transcription factors that can act as activators or repressors on the transcription of their target genes. In the terminology of molecular cell biology,

these transcription factors as well as the genes they are coded by are designated as trans-acting elements, while the regulatory DNA sequences where these transcription factors bind to regulate the target gene expression are called cis-acting elements. The trans-acting elements regulate gene expression by binding to the DNA at promoter-proximal elements (PPEs within 100-200 bps upstream from start site) or distant enhancer sites which can lie up to 50 000 bps (50 kbps, e.g. in SV40 enhancer) away from the start site. While the regulating mechanisms of enhancers consist in the formation of an enhanceosome through the assembly of several interacting proteins, there are several complex mechanisms by which the TFs binding to PPEs can control the transcription on the molecular level [18]:

- Activators can stimulate the assembly of the initiation complex
- Repressors can bind to a specific DNA site competitively to an activator or a general TF or can prevent in a different way the function of an activator or the assembly of the initiation complex

Moreover, there is a group of transcription regulating proteins that are not termed TFs because they don't have a DNA binding site. This group consists of chromatin remodeling proteins, histone acetylases, deacetylases, kinases and histone or DNA methylases [23, 24]. In this group, transcriptional activation or repression is mediated by chromatin remodeling through histone post-translational modifications [25] as will be explained in more detail in the following section.

### Regulation through epigenetic modifications

As mentioned in 1.4.1, DNA is associated with histone proteins in the nucleus to form chromatin. This highly condensed structure is formed of repeating units called nucleosomes which comprise approximately 146-147 bps of DNA wrapped around a histone octamer [26]. When folded into higher order structures, the repeating nucleosomes forming the chromatin are thus able to package approximately 2 m of DNA into a nucleus of roughly 10 µm diameter. The histones of the chromatin can be modified by specific enzymes in multiple ways which will affect the transcription activity of the part of DNA associated with them. Because to date, not all epigenetic modifications have been studied in the context of stem cell research, we will mainly focus on the modifications known to be associated with pluripotency, i.e. histone (de)methylation and (de)acetylation and DNA methylation.

There are two characterized mechanisms likely to occur as a consequence of histone modifications: first, the disruption of contacts between nucleosomes resulting in the unraveling of chromatin and second, the recruitment of enzymes able to modify the chromatin structure. Thus both

mechanisms are aimed at the remodeling of chromatin structure into either the transcriptionally active euchromatin or the inactive heterochromatin.

Considering the first mechanism, modification of the acetylation pattern has the highest potential to disturb chromatin structure, because it neutralizes the charges of the lysine residues it applies to, which may have important implications in intra- and intermolecular interactions of histones, especially their N-terminal tails [27, 28, 29].

For the mechanism of enzyme recruitment, recent findings suggest that the methylation pattern of histone H3 at lysine residues 4 (H3K4), 9 (H3K9) and 27 (H3K27) are important for the tethering of enzymatic activity including ATPases [30], histone deacetylases [31], demethylases [32, 33], ubiquitin ligases or methyltransferases [25].

Keeping the focus of this work in mind, it is important to assess the different modifications to one event, silencing or transcriptional activation of a gene. Recently, there has been evidence that acetylation of histones H3 and H4 as well as tri-methylation of H3K4 are marks of active chromatin whereas methylation of H3K9 and H3K27 and any deacetylation account for silent chromatin [25]. Nevertheless, as it is always the case in biological systems, there is not only one truth and given modifications may have the potential to act either in a repressive or an activating way depending on their location in the genome. H3K36 and H3K9 have lately been shown to act in such an opposite manner when found either in coding regions or in a promoter of a gene [34]. Moreover, activating and repressing chromatin signatures can coexist in so called bivalent domains in ESCs [35, 36] as will be depicted in 2.2.2.

### The role of miRNAs in transcriptional regulation

In addition to TF and epigenetic control, I would like to mention the recently discovered regulation of genes by small non-coding RNAs (ncRNAs) especially micro RNAs (miRNAs) which seem to play an important role in ESC pluripotency. Although the first miRNA was already discovered in 1993 in the worm C.elegans [37], the name of micro RNAs only came up in 2001 [38]. The miRNA family consists of 21-25 nucleotide small RNAs which are transcribed from the genome and processed to the mature miRNAs that negatively regulate gene-expression post-transcriptionally [39, 40, 41]. The short RNAs form imperfect stem-loop structures and bind to the 3' untranslated region (3' UTR) of their target mRNAs either acting as translational repressors [42] or inducing the degradation of the mRNA [43] depending on their complementarity. Important miRNAs involved in ESC pluripotency, as will be discussed below, are let-7 which seems to form a feedback loop with *LIN28* [44] one of the mentioned reprogramming factors of iPSCs [6], miR-21 which is involved in the regulation of *NANOG*, *SOX2* and *POU5F1* [45]

and miR-145 which regulates the expression of *POU5F1, SOX2* and *KLF4* [46]. Thus, these miRNAs can be integrated in the transcriptional regulatory networks involved in the iPSC reprogramming process where they may be playing important roles as repressive switches (see 4.3).

So far, we have mentioned some of the most important mechanisms of eukaryotic gene control that are involved in the complex regulatory mechanisms that control pluripotency and self-renewal. In the next section we will introduce the concept of gene regulatory networks (GRNs) which are crucial for the modeling of the processes associated with differentiation and reprogramming.

### 2.1.3. Gene regulatory networks (GRNs)

The genome of an organism plays a central role in all kinds of cellular processes. A GRN is a selection of certain genes interacting through mutual activation or inhibition by their encoded gene products. Every organism, be it a single-celled bacterium or a highly developed animal consisting of trillions of cells, needs to respond adequately to external conditions and signals in order to survive or to optimize its situation. Other than the direct and fast responses through receptor-protein coupled signaling, in many cases there is also a need for adjustment of the protein composition of the cell for a longer time range. This can be attained by the cell through regulation of the expression rate of its genes. A classic example is the adjustment of metabolic pathway's enzyme concentrations as a response to external substrate abundance. A lot of work has been done on this subject for metabolic and signaling pathways in prokaryotic model systems such as the lac-operon in *E.coli* [47] or eukaryotic cells, e.g. the cell cycle in yeast or *Saccharomyces cerevisiae* [48]. Other prominent examples of GRNs are the signaling pathways such as TGFß, MAPK, WNT and many more which will be discussed later in this work.

Since nearly all cells of an organism contain the same genetic material, it seems necessary to understand which genes are expressed to which extent at which time. Even though proteins involved in gene expression and molecular mechanisms of gene regulation is a well studied research field, the knowledge of the functioning of regulatory systems including these genes and proteins still needs to be expanded.

In the graphical representation of gene regulatory networks, we often don't differentiate between genes, mRNAs and proteins, especially when it comes to the mathematical modeling as will be explained in 2.3.2. Of course this can only be done if we accept, that we are not looking at an exact biological process which would involve transcription, processing, translation and maturation of the protein. To describe the pure interaction between the species we will identify

the nodes in the GRN, i.e. the species, with the expression level of a gene. Like that, in a logical network, if a gene is expressed and the corresponding TF represses another gene, than the ladder will be deactivated. Thus, there is no need to incorporate the transcription factor as a proper species when dealing with gene expression. Genes, or more exactly their expression level, are thus represented as nodes in a graphical representation of GRNs and interactions between the genes as edges. As there are two kinds of interactions, positive or negative, there are also two types of edges, i.e. activation is commonly represented by a normal arrowhead while inhibition is represented by a different arrow with a crossed out point.

The enormous interconnectivity of GRNs involved in physiological processes guarantees the perfect function of the organism, i.e. the adjusted answer to a diversity of possible inputs, but it also renders the analysis and dynamic modeling of such GRNs a challenging task. In fact, GRNs easily give rise to multiple dynamic behaviors even if the network structure still seems quite simple. The aim of every mathematical model is to understand how a GRN dynamically responds to diverse inputs, i.e. what changes in the expression of which genes will be effected. When confronted with large GRNs, a common approach to begin the network's study is the discrete Boolean modeling method which will be depicted in 2.3.1. That method may give a first insight into the behavior of the system before starting a lot more complex mathematical approaches such as qualitative standardized dynamical models or ODE systems 2.3.2.

## 2.2. Biochemical characteristics of ESCs

### 2.2.1. The core network of pluripotency: an example for a simple GRN

As seen in the last section, the dynamic behavior of GRNs is responsible for the function of many physiological processes. In this section, we will discuss how pluripotency is determined and maintained by the action of a core GRN which acts on a large part of the genome.

To understand human development and the therapeutic potential of ESCs, it is fundamental to identify the most important regulatory factors that determine cell fate. The transcription factors OCT4 (encoded by the gene *POU5F1*), SOX2 and NANOG were recently found to play essential roles in early development and ESC identity in mouse and human [49, 50, 51, 52, 53, 54]. Silencing any of these 3 factors in ESCs resulted in the differentiation of ESCs into one of the cell lineages of the 3 embryonic germ layers, thus showing their importance for the maintenance of pluripotency and self-renewal [55, 56, 57, 58, 59].

In 2005, Boyer et al. in a groundbreaking work identified *OCT4*, *SOX2* and *NANOG* as parts of a core transcriptional regulatory circuitry in hESCs. They carried out a nearly genome-wide ChIP-

on-chip analysis, a procedure that aims at revealing all the DNA binding sites of the transcription factors in question. Their results showed that the 3 transcription factors are master regulators, i.e. they bind to a large number of genes (around 600 for OCT4, more than 1000 for NANOG and SOX2) and that many of these target genes are co-occupied by a combination of at least two of the 3 TFs [60]. Some of the target genes were found to be active, e.g. those involved in pluripotency, others to be inactive, e.g. those having developmental functions, suggesting the master regulators to have activating and repressive functions at once depending on possibly involved cofactors, precise concentration levels or posttranslational modifications. Moreover, in the same publication, Boyer et al. postulated a regulatory circuitry consisting of interconnected autoregulatory loops between the 3 TFs.

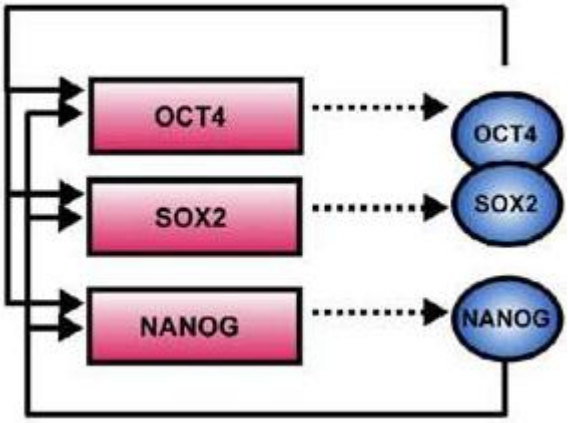


Figure 2.2.: The core regulatory network of pluripotency (from [60])

As one can see in Figure 2.2, the TFs OCT4 and SOX2 act positively together as a heterodimer on the transcription of their own genes and on that of *NANOG,* as postulated as well in other publications [61, 56, 62], while the TF NANOG acts positively on the expression of its own gene, *POU5F1* and *SOX2* [63, 60]. These findings were confirmed and expanded by many other groups later on to give more and more complex pluripotency networks [64, 65, 66, 67, 68] some parts of which will be mentioned later on in this work again.

We will now concentrate on the three transcription factors and their function in the ESC state maintenance. The transcription factor OCT4 which is encoded by the gene *POU5F1* is a pivotal regulator of pluripotency. Its artificial repression in ESCs leads to differentiation into the trophectodermal lineage. In the ESC state, OCT4 associates with CDX2, a trigger for trophectoderm (TE) differentiation, to form a complex that represses *POU5F1* as well as *CDX2* expression. As the TFs OCT4 as well as CDX2 regulate their own expression positively, down-

regulation of *POU5F1* will result in up-regulation of *CDX2* and vice versa [69] thus leading to a bistable system where either of the 2 factors is present at high levels and the other is not. However, it was shown, that if *POU5F1* is overexpressed in ESCs they will differentiate into primitive endoderm cells [70, 59] showing that a precise regulation of OCT4 levels in ESCs is required for the maintenance of this state. Moreover, looking at the great variety of its downstream target genes, OCT4 is very likely to be involved in a number of different cellular processes that have an impact on cell fate. A good overview of the processes, OCT4 and also SOX2 may be involved in is given in Greber et al 2007 suggesting multiple action on metabolism, cell cycle, chromatin architecture, signaling pathways or apoptosis only to name a few [71].

The TF SOX2 whose depletion in ESCs leads to defective epiblasts and differentiation into trophectoderm cells [57] is another part of the pluripotency core network. *SOX2* deficient cells die immediately after implantation while knockdown of SOX2 in ESCs induces differentiation into multiple lineages. We've already mentioned that SOX2 acts together with OCT4 on a large number of downstream targets activating or repressing their transcription. However, one main difference between SOX2 and OCT4 is that SOX2 is not restricted to ESCs but it is also present in extraembryonic ectoderm and neuroectoderm [57]. Nevertheless, similarly to OCT4, small increases in Sox2 levels in mESCs and embryonic carcinoma cells (ECCs) were recently shown to lead to differentiation into several cell types but not into endoderm [72]. In fact, when slightly overexpressed, Sox2 decreases the expression of its own gene and of *Pou5f1* and *Nanog* mESCs and ECCs [73]. Moreover, while OCT4 currently appears to be irreplaceable for reprogramming, *Sox2* could be substituted by a combination of *Sox1* and *Sox3* [74]. Overall it seems that the known role of SOX2 mainly consists in its action as a cofactor of OCT4 but as its exact expression level is critical for pluripotency and proper differentiation of reprogrammed iPSCs [75], it may have a more powerful task which still needs to be examined.

The third TF of the core network is NANOG which is highly expressed in the cells of the ICM and later restricted to pluripotent cells. Similar to the other factors its lack in mouse embryos results in the failure to develop an epiblast, i.e. an embryo, while NANOG knockdown in ESCs drives them to differentiate into extra-embryonic endoderm cells [76, 67, 51]. However, unlike *POU5F1* and *SOX2*, *NANOG* overexpression does not induce differentiation but promotes self-renewal and the maintenance of the pluripotent state [49, 51]. Apart from endoderm differentiation NANOG was also shown to block neuronal [77] and mesodermal [78] differentiation. There are many theories about the multiple functional mechanisms of NANOG. Thus, it was hypothesized that it acts as a repressor on *GATA6*, a gene coding for a transcription factor that triggers differentiation into primitive endoderm [79, 51] while Ralston et al. 2005 and Chickarmane et al. 2008 even suggest a mutual inhibition between the two factors [15, 68].

To summarize, *POU5F1*, *SOX2* and *NANOG* form a highly interconnected GRN whose overall task is the transcriptional activation of genes involved in pluripotency and self-renewal and the repression of genes responsible for differentiation processes. Nevertheless, recent work on the subject of up- or downregulation of the core TFs has proved that this simplistic view of the core network should be further extended to explain the narrow expression limits in ESCs. Thus, even though at the big level, a picture of ESC pluripotency and self-renewal is starting to arise, details of the numerous mechanisms and their importance for ESCs and iPSCs are still only poorly understood. We will try to give a short overview of the epigenetic modifications and signaling pathways involved in ESC maintenance and differentiation in the following section.

### 2.2.2. Epigenetic modifications in ESCs

In the last section we have discussed the pure core GRN responsible for the maintenance of the ESC state consisting of OCT4, SOX2 and NANOG. In this part we aim at describing the downstream effects of this network when it is switched on or off and possible effectors that target the genes of the core network.

Epigenetic processes that have been described in general in 2.1.2 were recently found to be involved in important mechanisms of cell fate determination and cellular reprogramming [80, 81, 82]. In fact, ESCs are overall transcriptionally more active than differentiated cells. Thus, during differentiation of ESCs to developmentally more restricted states (or vice versa) the global genetic expression patterns change due to covalent histone modifications, DNA methylation and localization of chromatin to specialized nuclear domains. The exact mechanisms and processes will be outlined in the following.

As mentioned above, gene activity critically depends on the chromatin structure which can be altered through histone modifications. When inducing differentiation in ESCs, e.g. by retinoic acid (RA), an increase of H3K9me3 (a heterochromatin marker) as well as a decrease of global concentrations of acetylated H3 and H4 (euchromatin markers) were observed [83, 84] showing that cells tend to develop transcriptionally less active chromatin.

We've already introduced the concept of bivalent domains in which active and inactive chromatin marks can coexist in particular regions mainly in ESCs. These domains, consisting of a combination of the transcriptionally inactive marker H3K27me3 and the transcriptionally active marker H3K4me3, are associated with genes expressed at low levels [35, 60] which is a common feature of many genes that relate to the ESC state [85, 86]. Since they are mainly found in ESCs and only scarcely in differentiated cells where either activating or inhibiting chromatin marks are present (but not both at a time), Bernstein et al. 2006 proposed these bivalent domain's main

task to be the silencing of developmental genes in ESCs while keeping them poised for activation when differentiation programs are initiated. Furthermore, they showed that upon differentiation bivalent domains resolve to either H3K27me3 or H3K4me3 resulting in a transcriptional inhibition or activation of the associated gene respectively. In fact, it has been found that many target sites of OCT4, SOX2 and NANOG co-localize with such bivalent domains, these genes tending to be silenced in the ESC state [35]. Implications of these mechanisms in transcriptional models for ESC pluripotency will be hypothesized below in the epigenetic model of the result section.

DNA methylation at CpG islands is another gene silencing mechanism that has not yet been mentioned and that is required for the induction of differentiation of ESCs. In this context, hypomethylation of the promoter of pluripotency related genes is required for high expression levels and thus for maintaining pluripotency. Genes involved in these processes include the DNA methyltransferases Dnmt3a and Dnmt3b and the CpG binding protein (CGBP)[87, 88].

### 2.2.3. Signaling pathways involved in pluripotency

Pluripotency of mouse and human ESCs is determined by different signaling pathways. In mouse ESCs (mESCs) one key factor to maintain pluripotency and the ability to self-renew is the leukemia inhibitory factor (LIF) [89, 90] which works by activation of the Janus-associated tyrosine kinase (JAK) and of signal transducers and activators of transcription (STAT) [91, 92] which activate downstream targets involved in the maintenance of the ESC state such as the key factor c-Myc [93] that is able to induce infinite proliferation through the control of telomerase [94]. While it is known that STAT3 is the main mediator of LIF signaling, the exact mechanism remains largely unknown. Nevertheless it has been shown that bone morphogenetic proteins (BMPs) are likely to play a role in LIF signaling as well. In mESCs in fact, depending on the presence of LIF, BMPs either induce mesodermal differentiation in the absence of LIF or help promote self-renewal when LIF is available [77]. However, in human ESCs (hESCs) the ESC state has been found to be independent of this LIF-STAT3 interaction keeping its property in vitro [95].

Another difference between mESCs and hESCs was found in the mitogen-activated protein kinase (MAPK) signaling pathway and especially the extracellular-signal regulated kinase (ERK) which belongs to the MAPK family. While mESCs show high activity of extracellular-signal regulated kinases (ERKs) when inducing differentiation and ERK suppression leads to self-renewal [96], in hESCs high levels of ERK activity have been found, relating to basic fibroblast growth factor (bFGF) signaling which can maintain hESCs self-renewal over a long period of time [97, 98, 99]. In fact, hESCs are most commonly cultured on feeder layers of fibroblasts in

the presence of bFGF which seems to have great importance for hESCs self-renewal [100, 101]. Components of the bFGF pathway have thus been found to be enriched in hESCs and inhibitors of FGFR for example lead to their differentiation [102]. It has moreover been discovered recently that hESCs cultured without feeder layers but supplemented with bFGF and Noggin (an inhibitor of BMP4) are supported in their undifferentiated proliferation [103, 98, 104] although very high concentrations of bFGF (100 ng/ml) have also been proved to be sufficient for the maintenance of the hESC state [105]. The mechanism by which bFGF can help maintain pluripotency and self-renewal are largely unknown though it was recently found that it can activate MAPK pathway [98, 97] and inhibit BMP4 activity on the SMAD1 level [103].

Another factor that has an effect on the ESC state is the TGF beta pathway family which comprises more than 40 members including BMPs, Activin, Nodal and TGF beta itself. We've already mentioned BMPs above where we showed that their function in mESCs is dependent on the presence of LIF. In fact, BMP4 acts through phosphorylation of SMAD1/5/8 whose activation leads to the expression of the inhibitor of differentiation (Id) that blocks neural differentiation [77]. It has also recently been found that SMAD1/5/8 directly acts on *SOX2* decreasing its expression [106]. Moreover BMP4 blocks the MAPK cascade maintaining the pluripotent state in mESCs in the presence of LIF [107]. In contrast to its effect in mESCs, in hESCs BMP4 does not maintain the ESC state but induces differentiation to the trophoblast or primitive endoderm lineage [108]. The other branch of the TGF beta pathway that leads to the activation of SMAD2/3 can be activated by TGF beta, Activin or Nodal. In this case SMAD2/3 inhibits SMAD1/5/8 showing that the TGF beta/Activin/Nodal pathway can block BMP4 mediated signaling [109, 110]. Thus, it was shown that without condition medium, Activin A can keep hESCs in an undifferentiated state [109]. Moreover Greber et al. showed in 2007 that SMAD2/3 directly enhances *NANOG* expression [111]. These interactions will be the basis for our pathway model in 4.4.

To summarize, Figure 2.3 shows the involvement of the different pathways and their consequences for the ESC pluripotent state. Overall, these relations of the signaling pathways JAK-STAT, MAPK, FGF, TGF beta suggest that they play important roles in ESCs self-renewal and pluripotency as will be reflected in our models in the results section.

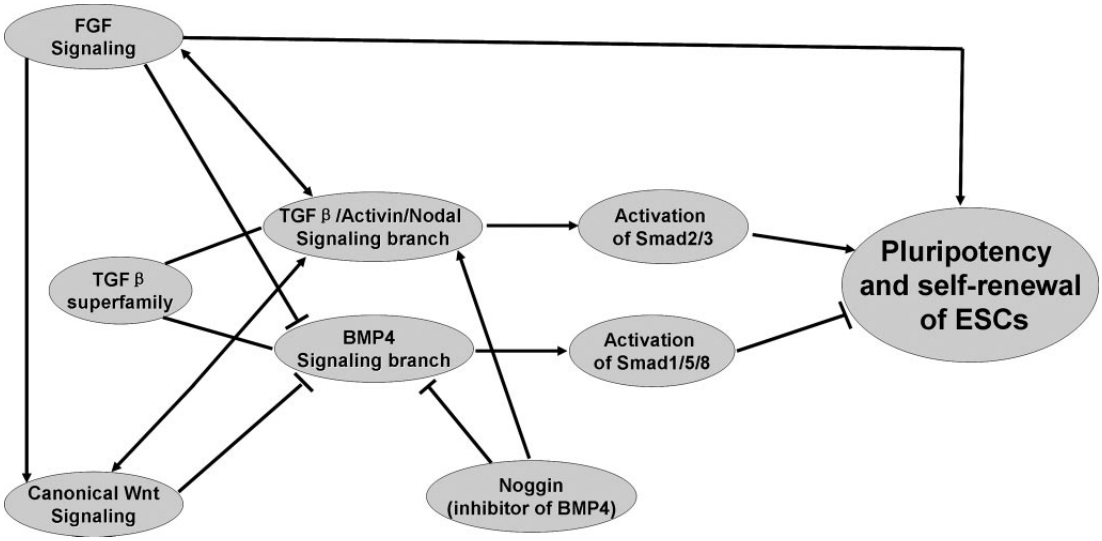


Figure 2.3.: Overview of the relationship between the relevant pathways and their implications for the ESC state

### 2.2.4. The task of miRNAs in the ESC fate

As seen in 2.1.2, where we described the mechanisms of eukaryotic gene control, several miRNAs have been identified as active parts of the transcriptional network regulating pluripotency and self-renewal. As they are going to play an important role in the modeling of regulatory networks responsible for ESC fate in the result section, we will focus on miR-21 and miR-145 and their relations to the pluripotency core network.

It has recently been shown by Xu et al. that the microRNA miR-145 is expressed at low levels in hESCs but strongly upregulated during differentiation. Moreover, they found out that the genes of the core transcription factors OCT4, SOX2 and KLF4 are targets of miR-145 which inhibits their expression. However there seems to exist a negative feedback loop on the transcription of miR-145 by OCT4 thereby creating a network of interconnected inhibitions [46].

Another miRNA involved in the determination of pluripotency and self-renewal in mESCs is miR-21 which acts as an inhibitor on the genes of the core TFs OCT4, SOX2 and NANOG. In the undifferentiated state, this miRNA is repressed by the neuronal repressor protein RE1-silencing transcription factor (REST) [45] that restricts the differentiation along the neuronal lineage [112, 113]. In fact, REST is directly activated by OCT4, SOX2 and NANOG [60] in the ESC state, thereby keeping miR-21 repressed which shows once more the autoregulatory mechanism of the core network of pluripotency that keeps itself activated.

We have shown that the miRNAs -21 and -145 are actively involved in the fate decisions of stem cells repressing the expression of core TFs when they get activated by an external or internal signal, e.g. upon differentiation when the core TFs are downregulated. In the result section of this work, we will see which role miRNAs can play in the dynamic modeling of pluripotency networks.

## 2.3. Methods for the mathematical modeling of GRNs: A systems biology approach

The main task of the systems biology approach is to explain the dynamic behavior of biological systems observed experimentally. In this definition the biological system in question can be any set of compounds, i.e. genes, proteins, mRNA, small molecules, etc. inside or outside of a cell that can be translated into a set of variables describing their respective amount or concentration. Here, we will concentrate on the mathematical modeling of GRNs, i.e. the variables will reflect the concentration of a specific mRNA or a protein because in GRNs we don't discriminate between these two species. The evolution of each species' concentration over time will then be defined by certain rules depending on the chosen mathematical approach.

There are several mathematical methods that were shown to be able to describe cellular, molecular or physiological processes with high precision, i.e. reflecting experimental data in a good way (examples: cell cycle in yeast, glycolysis, other pathways in other organisms (look for examples, citations, etc.). These methods which will be presented as follows divide into two major fields, one applying binary logic to a system of variables and the other using ordinary differential equations (ODEs).

### 2.3.1. The Boolean modeling approach

The foundations of Boolean logic were laid by George Boole in his work *The mathematical analysis of logic* which was published in 1854. His ideas were then further developed by Jevons, Peirce, Huntington, Stone and others over the last century to the Boolean algebra that we know today. We will herein directly focus on the application of Boolean logic to biological networks which was first described by Stuart Kauffman in the 1960s [114].

As we have seen before, GRNs can be represented in directed graphs where the nodes correspond to the genes and transcription factors in question and the edges describe the effect, i.e. transcriptional activation or inhibition of the input node on a target node. In the Boolean method

all the nodes are assigned Boolean variables that can only take two values, 0 or 1. That is to say each gene represented by a node is either active or inactive at a certain time t like a switch that has only these two positions. This assumption is legitimate as many important reactions in vivo are nearly irreversible and follow a sigmoidal rate dependence on the concentration of the metabolites both in enzyme [115] and gene control systems [116]. The ensemble of all the variables' values at a time t is called the state of the network at the time t. As every variable can take two values, in a network of n nodes there will be $2^n$ possible states.

The progression of all the variables in time and thereby of the state is determined by rules, the so called Boolean functions, that depend on the structure of the genetic network. In that context we will call "input node to a target node" every node in the graph of the network that is connected with the target node through an edge that points towards the target node. For every node, the effect of all its input nodes is described by a Boolean function that will determine the following state t+1 from an initial state t in a deterministic way. A Boolean function $f_i$ for the node i takes as an input all the values of the input nodes to the node in question at time t and generates an output value for the node i at time t+1. Boolean functions can easily be represented by truth tables as it is visible in Tab.2.1 where the OR and the AND function for 2 inputs are represented. Every input has 2 possible values 0 or 1 so for every Boolean function there are $2^2$= 4 outcomes that have to be defined. Every of these outcomes can again take two possible values, hence there are $2^{2*2}$= 16 possible combinations of outcomes, i.e. Boolean functions for this specific problem or more generally for k inputs $2^{2k}$ possible Boolean functions [114].

| OR | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 1 |

| AND | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

Table 2.1.: The truth tables for the Boolean functions OR and AND

Hence, time step by time step, the system develops to finally reach a state, that it has already visited before. Since the dynamics are deterministic the system will fall into an attractor, i.e. a steady state or a stable oscillator.

The biggest challenge in this Boolean modeling method is to find the right Boolean function for every node in the network. There are two main ways to deal with this problem. One can either try to construct random nets and have a look at their dynamics comparing them with the real biological system like Kauffman did in several studies [114, 117, 118] or one can choose to define specific Boolean functions for each node when there is sufficient experimental information on the interaction between the genes. We will chose to go with the second step since in most of our models we at least have the interaction information of the GRN. The implementation rules

for the conversion of the directed graph representing the GRN into the Boolean model will be mentioned in 3.1.

The state space of a Boolean network is the ensemble of all possible states and their progression in time. It can be represented in a directed graph where every node corresponds to one state and the directed edges mark the transition from one to the next state in time. A sequence of transitions between states is called a trajectory or a path. Such a graphical representation of the state space of a Boolean network contains a wealth of information in a very easy appearance. First of all, we can deduce the steady states and possible stable oscillators, i.e. the attractors of the system that the variables will reach after a certain time. A state in the graphical representation of the state space can be called steady state if there are no outgoing edges from it, only incoming edges towards it, meaning that this state remains as it is over time and does not change anymore. Meanwhile, stable oscillators can be detected by finding cycles in the graph, i.e. closed paths where the start and end node are the same. Moreover, although a state space seems to be a static representation more than a dynamic model, we can deduce the dynamics of the network by looking at the trajectories, i.e. at the transitions between the states. When looking at the trajectories through the state space, it also gives us information about the reachability of specific states when starting from a certain different state. In fact, a state can only be reached from another one if there exists a path or trajectory leading from one to the other either directly or through different intermediate nodes. Thus, the state space is an easy graphical representation that can provide us with a lot of information about the network stability, its dynamics and reachability of certain states. For a better understanding, we will give an example for the graphical representation of the state space of a Boolean model and its contained information below in 3.2 where we also mention the software used.

There are two main forms of dynamic Boolean modeling that differ from each other in their application and prediction power, the synchronous and the asynchronous approach. While the update rules, i.e. the Boolean functions, are applied simultaneously on all the variables of the system in the former, in the ladder they are applied in a random order thereby introducing the concept of stochasticity into the Boolean method. In fact, when repeating the asynchronous approach several times, we will find an ensemble of possible steady states or state cycles each of which will be attained with a certain frequency. This frequency can be extrapolated to a probability when comparing to the total of runs.

For further reading on the Boolean method and its application to biological systems, I suggest the works of Kauffman [114, 117, 118] and the reviews of Huang and de Jong [119, 120].

The Boolean modeling approach is an example for a discrete model, i.e. the variables in the network can only take a countable number of values, in the Boolean example only two. These

discrete methods have recently served to model and predict the main characteristics of complex networks and large-scale regulatory systems [121, 122]. In 2006, Mendoza and Xenarios presented a method for the extrapolation of any regulatory network graph to a continuous dynamical system. It is based on the construction of two dynamical systems one discrete, one continuous from the network structure. The stable steady states of the discrete system can then be found analytically and used to find the stable states of the continuous system numerically [123]. Together with Di Cara, they moreover created their own software, SQUAD that is able to perform this extrapolation. This software and its technique will be further described in the result section when we will focus on the mathematical analysis of different versions of pluripotency related GRNs [124].

A last concept that has to be introduced is the one of the hamming distances. A hamming distance between two states of a Boolean network's state space is defined as the minimal number of variable values that have to be changed to reach a certain state from the other one [125]. We used this concept in the result section to unravel possible transitions between subnetworks in the state space thus being able to manipulate a transitional state that would normally end up in a certain steady state to join another subnetwork that reaches a different steady state. This concept picks up the idea of reprogramming by manipulating a terminally differentiated state or any transitional state that would end up in this differentiated state and only introduce a minor change (e.g. one gene activation or inhibition for hamming distances of 1) in this state to end up in a pluripotency related steady state or a transitional that reaches this steady state.

Although Boolean networks do not correctly model every biological process, e.g. the dynamics of a TF that downregulates its own expression as shown in [118], and their state space increases exponentially with the size of the network, they nevertheless constitute a well suited model system for small networks when only qualitative data is available. In fact, the Boolean method always provides a good insight into the existence and stability of steady states and the robustness of the network.

### 2.3.2. Modeling of GRNs by ordinary differential equations (ODEs)

Ordinary differential equations (ODEs) are the most widespread formalism to model dynamic systems not only in biological but as well in chemical and physical systems. Strongly applied to GRNs, this formalism describes the time-dependent variation of RNA, protein and other molecule concentrations with variables that are connected through rate equations in a differential way. The system with n variables usually has the following mathematical form:

$$\frac{dx_i}{dt} = f_i(x_1, ..., x_n)$$

where $x_i$ is the $i^{th}$ species' concentration and $f_i$ is the rate equation, usually a non linear function that has to be defined, connecting $x_i$ with the components of the systems which it is interacting with, i.e. in a GRN its activators and inhibitors. The rate equations describe the production rate of the species and can either be chosen or combined out of a set of different types of equations, e.g. mass-action kinetics, Michaelis-Menten kinetics, Hill kinetics or they can be derived by the exact mathematical description of the network. When enumerating the equations for all the species, we are facing a system of non-linear ordinary differential equations describing the dynamics of a GRN.[120]

To model a GRN in a detailed, biologically relevant model, one would have to take into account all the steps from the gene to the protein as largely described in 2.1. This would mean that genes, mRNAs, proteins and to be really exact all the intermediate steps of the process would have to be proper species which would increase the number of variables of the model very quickly making it very complex and unclear. To overcome this problem, we can simplify things and assess the variables to the activity of the genes which can be increased by synthesis and decreased by degradation of the species in question. As well as the example employed in [126] which is shown in Figure 2.4, we also made the assumption in our models that every transcript is immediately translated into protein, the kinetic constants of synthesis thus referring to transcription and translation.

We will concentrate now on the derivation of the kinetic rate laws for the different interactions in a GRN that define the ODE system. As mentioned above, we mainly use Michaelis-Menten (MM) or in rare cases Hill kinetics to model the interactions. When dealing with a single inhibition as shown in Figure 2.4a) in the first equation or with a single activation as in the second equation, we just admit the MM or Hill kinetics with the only difference that we use the form

$$\frac{d[gene1]}{dt} = \frac{V1}{k1 + [gene2]^n}$$

for inhibition and

$$\frac{d[gene1]}{dt} = \frac{V1 * [gene2]^n}{k1 + [gene2]^n}$$

for activation, where most of the time we admit n = 1 which leads from general Hill kinetics to MM kinetics. It is moreover possible to include basal transcription rates, i.e. constants that will be added to the equation.

Now, if we have to deal not only with one but with more inputs to a gene, i.e. 2 activations or an activation and an inhibition, we combine the equations for the two following the upcoming rules. For several activations we just sum up the terms for the activation which leads us to an increased overall response when all genes are active but still preserves the ability for the gene

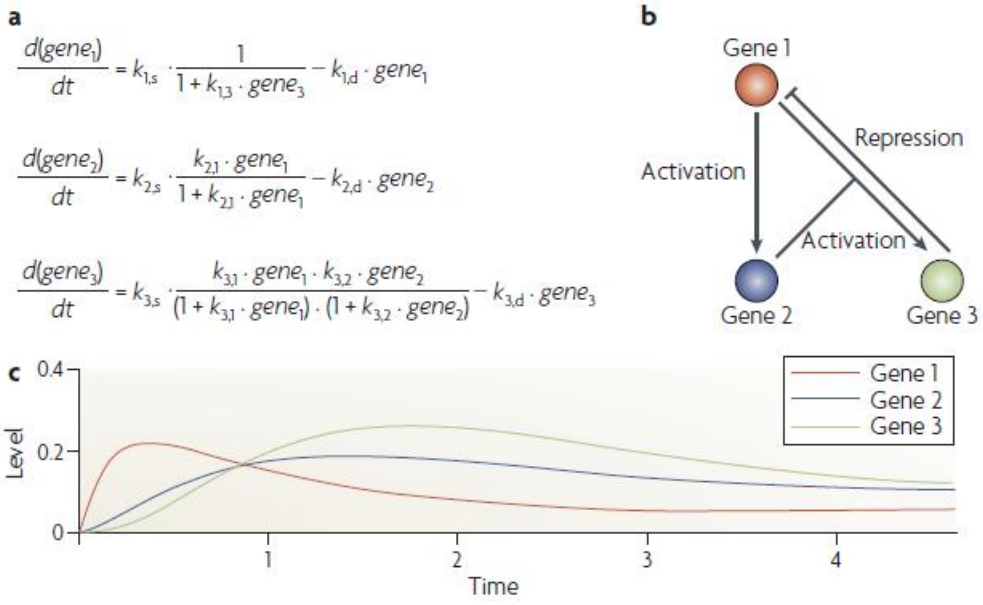


Figure 2.4.: An example for an ODE model from Karlebach and Shamir 2008 [126] A network of three genes shown in **b** is modeled by the system of ODEs shown in **a** Every differential equation is assembled by a term for synthesis and one for degradation. The term for synthesis depends on the activators or inhibitors of the gene in question. The kinetics used for activation in here are Michaelis-Menten Kinetics, for repression it is for degradation it is simple linear (mass action) kinetics, The graphical representation in **c** shows the expression level of the respective genes over time, thus the dynamic behavior of the system.

to be active, when not all activators are present. To include an inhibition into the equation with one or more activations we will just multiply the denominator of each activation term with the denominator of the inhibition equation in the way that every activation is lowered by the presence of inhibitor. We can compare this with the OR and the AND gates of the Boolean modeling: for several activations, we just used an OR connection, which corresponds to the sum in the ODE model, while for an inhibition in combination with an activation we always used the AND gate combined with the NOT operation for the inhibitor which corresponds to the multiplication in the denominator in the ODE model.

The ODE approach provides detailed information about the network dynamics but when it comes to fitting the system to the biological reality, it requires the presence of a wealth of high quality data for the estimation of kinetic parameters which often are not available.

# 3. Materials and methods

## 3.1. MS Excel, the simplest tool for dynamical Boolean modeling

When it comes to dynamic Boolean modeling, i.e. the representation and calculation of the time dependent behavior of an implemented Boolean network, software is scarce until today. I will give a quick overview of existing software and its problems in the following.

As a tool for matlab, CellNetAnalyzer is an interesting software to create and analyze Boolean networks that we tried to use in first place. It combines a wealth of possible actions and provides the possibility to fully analyze Boolean networks but it turned out to be hard to handle because of its complexity and different features were not well suited for our purposes. Another software dealing with the representation and analysis of Boolean networks, called GINsim (gene interaction network simulation), was found to be not trivial to handle and is probably better suited for larger networks.

Compared to these tools, MS Excel at first sight does not seem like the best choice to perform an analysis of Boolean networks. However the huge advantages of this software consist in the fact that one knows exactly what it does, that Boolean functions can easily be implemented and one can qualitatively perform an iterative time course that adjusts itself automatically when the input values are changed. As everyone probably knows how to deal with MS Excel and as the method is really straight forward, its following description will be short and comprehensible.

As a first step, we put the names of our species in the first row and leave the second row empty in the first place. The ladder will serve as our input value row when the Boolean functions will have been defined. Now in the third row, we can just use MS Excel's logical language comprising not more than the 3 operators "And", "Or" and "Not" which you can build every Boolean function with, and connect the values of the first row in the way they need to be connected to reflect the GRN structure. The rules for this connection are the following:

1. Without any inputs, a node at time t+1 takes the same value that it had at time t. Such a node is called an input node to the network.

2. When affected by activations and inhibitions, the value of a node at time t+1 is determined by the following function:
$N(t+1) = (\underset{i}{OR}(A_i))AND(NOT(\underset{i}{OR}(I_i))$

where $N(t+1)$ is the value of the node at time t+1, $A_i$ is the value of the $i^{th}$ activator of the corresponding gene and $I_i$is the value of the $i^{th}$ inhibitor of this gene. The indices underneath the OR and AND functions indicate, that the activators (respectively the inhibitors) are all connected through the OR function. That function implies that one active inhibitor is enough to inhibit the expression of the gene in question even if many activators are expressed. This assumption is probably the most limiting one in the Boolean method but since we don't have any information on the subject of the molecular mechanisms of inhibition and activation of the genes and transcription factors in question in pluripotency related GRNs, this is at least a reasonable assumption to make. When looking at the rules one may ask where we include the degradation of the species. It is very easy since upon lack of an activator for a certain gene at time t, it will be turned off at time t+1 because in that case the Boolean function returns a 0.

The rules as they are defined above are applied exactly like this in MS Excel for every cell of the third row. In the next step, we can simulate an iterative time course by applying the rules progressively on the cells of the next rows just by clicking on the third row in the lower right and pulling the mouse straight down the next rows. We can stop this step-by-step time course, when the cells don't change anymore or when we discern a loop structure, i.e. when a complete row is only a repetition from an earlier row, since the update rules are deterministic.

To visualize the Boolean modeling method with MS Excel, we will now give an example with the pluripotency related GRN by Chickarmane and Peterson from 2008 [68]. The network, its implementation and the results for 4 different input value combinations are shown in Figure 3.1. In fact, as we will see below in the result section, we could have eliminated the OCT4-SOX2 dimer as a proper species, but since we're only looking at the dynamics and not at the whole state space, which would increase in size by the factor 2 when including the OCT4-SOX2 dimer as a species, we can leave it in the network for the upcoming analysis.

Figure 3.1 shows the graphical representation, the implementation and the Boolean time courses for some initial values for the GRN proposed by Chickarmane and Peterson 2008. We were able to reproduce some of the outcomes that were found in the paper by ODE modeling, e.g. the two steady states where either the pluripotency related TFs are activated while the others are not or the other way around. Moreover we can show that NANOG alone can push the network into the pluripotency related steady state even in the absence of OCT4 and SOX2 which was a prediction of an earlier publication by Chickarmane et al. in 2006 [127]. However, the Chickarmane paper mainly focuses on the biphasic response of the network in dependence of

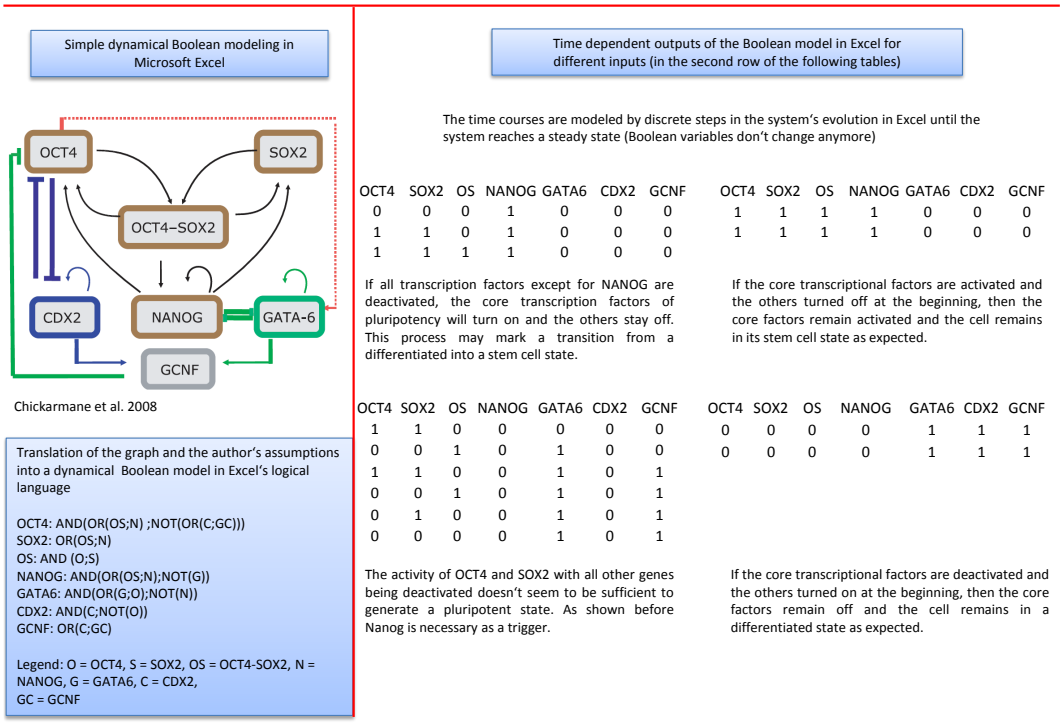


Figure 3.1.: The Boolean model of the network proposed by Chickarmane and Peterson in 2008 [68]. Upper left: The graphical representation of the GRN. Lower left: The implementation of the network in MS Excel's logical language. Right: Time courses with explanations for different initial conditions of the Boolean model.

the OCT4 concentration. In fact, the authors refer to earlier results of Matoba et al. [128] who showed that differentiation related genes can only be active at either high or low concentrations of OCT4 while pluripotency related target genes require intermediate OCT4 levels. Hence, Chickarmane and Peterson proposed a network to describe this bistable switch which describes different steady states for the 3 different levels of OCT4. As we decided to follow a Boolean approach for which the species can only take two values, it is impossible to reproduce these bistable switch data. To do so, we would have to use a multi-valued logical approach, work with Petri nets or ODE modeling. When implementing a multi-valued logical model, the definition of the update rules and functions gets more and more complicated with the number of values that a variable can take with the consequence that the implementation of an ODE model which can make better predictions is sometimes not more complex to do. However, as experimental data are scarce in this young field of stem cell research, we will stay with the Boolean logic in first place and give an outlook on an ODE modeling approach only at the end of this work.

## 3.2. BooleanNet and Cytoscape: a visualization of the state space for small Boolean models

In the last subsection we described the Boolean modeling approach with a simple tool, MS Excel, that was originally designed to create and work with data arranged in lists and tables. We will now outline the method we applied for the graphical representation of the state space of a Boolean network model illustrated by the same example we used above. Therefore we are going to introduce BooleanNet, a software toolbox based on matlab for the study of the dynamic behavior of Boolean models and Cytoscape, a software originally designed for the visualization of molecular interaction networks by graphs.

What mainly distinguishes BooleanNet from other efforts to model Boolean networks is its focus on well defined biological sub-systems rather than large-scale networks and its input simplicity. Similarly to MS Excel the input structure for BooleanNet is directly based on the 3 Boolean functions AND, OR and NOT which can build up any more complex Boolean function when combined. Once the update rules are implemented, BooleanNet can use synchronous or asynchronous iteration, stochastic updates or hybrid modeling by a system of linear differential equations. Moreover, which is not necessarily true for every tool that is designed for Boolean modeling, every aspect of the simulation can be altered throughout the process: node states can be overridden, updating rules changed and differential equations replaced [129].

Cytoscape is a visualization tool for the graphical representations of networks that supports many input formats e.g. SIF, GML, SBML, delimited text files and MS Excel workbook. In addition one has the possibility to import data files such as expression profiles or GO annotations which can be attributed to nodes, edges or whole networks which then can be manipulated, classified, arranged and structured based on this information.

To represent the state space of a Boolean network model, we first created a python script (not shown in this work) for BooleanNet which takes as an input the file containing the update rules in the BooleanNet format, the name of the variable species, the name and format of the output file, whether or not one wants the hamming distances file to be generated and whether or not the species should be arranged in different classes. The script generates 2 SIF files, one representing the whole state space of the model, the other one containing all the edges between nodes with an exact hamming distance of 1. Moreover it also creates property files for the nodes, containing whether they are steady states or not and another file representing their in-degree, i.e. the number of nodes they are a target of.

When loading the SIF files into Cytoscape, we get the state space as a directed graph that we can further manipulate. We can create different layouts of the state space for better comprehen-

sibility depending on the network's properties, i.e. steady states, classes, hamming edges, when importing these properties into the software. For better understanding, we will show here the same workflow we will perform on our own models in 4 for the Chickarmane model system of the last subsection. For simplification we have left out the OCT4-SOX2 dimer as a proper species as described above to reduce the number of states in the state space by the factor 2.

Its state space is shown in Figure 3.2. As explained in the graphic text, the state space is shape and color coded as it will be for most models in 4 as well. Thus, nodes with a spherical shape indicate transitional states while nodes with a triangular shape indicate steady states and the color indicate the state of the network from states that are more related to pluripotency (red) to states that are more related to differentiated states (green). When we take a closer look at Figure 3.2 we can discern 9 different subnetworks in the state space leading to 7 steady states and 2 state cycles. A subnetwork being defined as an ensemble of states leading to the same steady state or the same cyclic attractor.

Without going too much into the detail of this state space, the steady state 111000 (upper right) corresponds to the pluripotency related steady state. 000101 (upper left) and 000011 (middle right) respectively represent the differentiation related steady states where either GATA6 and GCNF or CDX2 and GCNF are activated. That is to say 000101 represents the endoderm and 000011 the trophectoderm state. Both of them can be attained by different initial or transitional states as can be seen by looking at the nodes directly or indirectly related with these steady states.

As is visible in the lower left and the lower right of Figure 3.2, when a subnetwork has no steady state, it has to contain a state cycle, i.e. a cyclic path of repeating nodes. In our example it's the state cycles 110001 <–> 011100 and 111100 <–> 110000 that are found in separate subnetworks that do not contain a static steady state.

Moreover, what we can do if we remember the dynamic simulation done in MS Excel in the last subsection, we can try to retrace the dynamics of certain initial values in MS Excel and in the Cytoscape state space. The only thing that we have to respect in this task is that we left out the OCT4-SOX2 dimer, so we have to compare the steady states without this extra species. We find that the two steady states found in MS Excel for the differentiation and the pluripotency related steady states are the same as in the state space. Moreover we can see that the steady state reached from the state where only NANOG is activated and all other genes turned off is the same in both examples even though the dynamics are different. This is due to the fact, that assumingly the OCT4-SOX2 species changes the dynamic behavior of the system. This is also why the dynamic and the steady state of the last MS Excel example for the initial values of OCT4 and SOX2 activated and the others turned off, differ from the ones in the BooleanNet/Cytoscape

state space. This effect will not be further explained in here but is due to the temporal shifting introduced by the activation of a dimer species (like in the section above) instead of a direct activation of a target gene by an AND gate like we did in here in the BooleanNet input file. We can learn from this example that we have to pay attention on the changes of the system's dynamics and stability when we introduce minor changes. Even if it seems like a trivial thing in our example that should not alter the dynamics profoundly, it nevertheless changed the whole state space.

Another concept that should be illustrated by a graphic is the concept of hamming distances mentioned above. To represent states that are closely related to each other, i.e. that could maybe be converted into each other by minor changes or by transcriptional noise, we introduced edges between all states, that only differ by the activation or inhibition of one species, i.e. two binary states that only differ in one digit.

Figure 3.3 shows the part of the state space from Figure 3.2 discussed above. We isolated the three subnetworks leading to the two steady states of differentiation and the steady state of pluripotency and only represented edges between nodes of different subnetworks that have a hamming distance of exactly 1. As we can see, transitions between the subnetworks leading to different steady states are possible from several nodes of the different subnetworks.

In fact, we now have analyzed the GRN from Chickarmane et al. 2008 with our Boolean modeling approach in MS Excel as well as with the combination of BooleanNet and Cytoscape which would make it justifiable to put it in the result section. However, as we think that the network constitutes a good example to understand how to work with these two methods, we let it in this section and just remind the reader that this actually is also part of the results we got out of the analysis of different networks involved in ESC pluripotency.

## 3.3. SQUAD: From discrete to continuous Boolean modeling, the standardized qualitative dynamical model

SQUAD is a software developed by Di Cara et al. in 2007 [124] for the purpose of dynamic simulations of signaling networks with the standardized qualitative dynamical systems approach which was developed by Mendoza and Xenarios in 2006 [123]. This approach can be used when kinetic and other experimental data are scarce or even not available at all. The software is able to convert the GRN in question into a discrete dynamical system in first place to identify its steady states. It then implements a continuous dynamical system applying defined rules that can be read in [123] and whose explanation would lead to far in this place. The steady states of the

continuous system are located near the steady states of the discrete dynamical system and can thus easily be found. The method thus ends up in an automatically generated continuous ODE system, which the software uses to generate time courses.

The input file for SQUAD only contains the network structure, i.e. the species and their activating or repressive interactions. The implementation is very straightforward: activations are defined by a "->" between two species, while inhibitions use a "-|". Thus, the input file for a network with a species A activating a species B and species B inhibiting species A would have the following form:

A -> B

B -| A

and this is everything that is written in the input file and passed to SQUAD.

For the simulation of time courses for different input values, we now have the possibility to adjust decay rates of the species, gains and attribute weights to the interactions between the species. The exact implications of these parameters in the continuous system can also be looked up in [123]. When all parameters are set, we can run time course simulations for different initial conditions and extract the dynamical behavior of the system as well as the steady state that is reached.

The big advantages of the software SQUAD are the easy handling, the scarcity of information that it needs in comparison to the wealth of information that it generates, i.e. we only tell SQUAD the network structure and it generates a continuous ODE model with adjustable parameters and time courses of the species' concentrations.

If we want to go a step further and develop our own ODE model with our own kinetic functions and parameters, the software to do so will be described in the following section.

## 3.4. Copasi: a COmplex PAthway SImulator for the integration of ODEs

In the last section, we discussed a transition from discrete Boolean to continuous modeling using an extrapolation method developed by Mendoza and Xenarios. This method automatically generates a set of ODEs with predefined rate equations and parameters which can be adjusted in the software SQUAD. However, this generalized logical approach is still only a qualitative dynamical method which is limited in its prediction possibilities and the analyst is only given a defined set of parameters he can adjust to fit his model to possible experimental data or biological

findings. Hence, the next step to model GRNs is the proper implementation of ODEs with kinetic rate laws and parameter sets that we develop on our own and that we think suit the problem best.

For the integration and analysis of our ODE models we used the software Copasi that stands for COmplex PAthway SImulator and is described in detail by Hoops et al. 2006 [130]. Copasi has multiple features including deterministic and stochastic simulation of reaction networks, stability, stoichiometric, sensitivity and metabolic control analysis, optimization and parameter estimation. In this work we only use the the integration of the ODEs for graphical representation of the time courses for different initial value and parameter sets. This is why, we will only describe these features of Copasi in slightly more detail and neglect the others.

The model can be implemented in Copasi by defining the reactions, the species and the form of the kinetic equations that can be chosen out of a library or defined properly, this information being translated by the software into a mathematical model, i.e. a system of non-linear ODEs. When specifying the parameters of the model and the initial values, Copasi can calculate the time courses. i.e. the development of the variables over time, using either a deterministic (LSODA integrator [131]) or a stochastic approach (Gibson-Bruck method of the Gillespie algorithm [132]) which can easily be switched by the user. The steady states of the system are determined by the damped Newton method and forward or backward integration using LSODA.

We will now give a short overview of how to implement a simple model in Copasi, how to adjust initial concentrations and parameters and how to run a simulation. In fact, to translate the graphical representation of a GRN into a reaction scheme, it is sufficient to build two reactions for every species, one of them characterizing its synthesis, the other one its degradation. While for the degradation we can assume linear mass-action kinetics because it only depends on the concentration of the species to be degraded, for the synthesis, i.e. transcription and translation, we have to include the activation and repression by other species. As derived in 2.3.2, we will declare the corresponding equations as rate laws for the reactions in question depending on the number of activators and inhibitors of the reaction. Following this step, we will still have to Figure out the parameters of the system to derive its steady states. This is probably the most limiting step of the ODE method, because we don't know the parameters of the reaction, since we don't have any quantitative data on molecular mechanisms. Nevertheless, we can try to have an insight into the qualitative behavior of the system by setting the parameters to random values (e.g. all to 1) in a first step and let the software run a time course. For quantitative information on the system and possible prediction, we will have to estimate the parameters which is only possible if we have sufficient data, i.e. time points for the concentration of the species in question, from experiments.

For the moment, we will stay with the qualitative behavior of the system. However, if we want to

look at specific time courses, we will have to think of a set of initial values that is appropriate to our problem. Hence, if for example, we would like to see what happens if all core transcriptional factors are active and the others are not, we will have to adjust our initial values depending on these conditions.

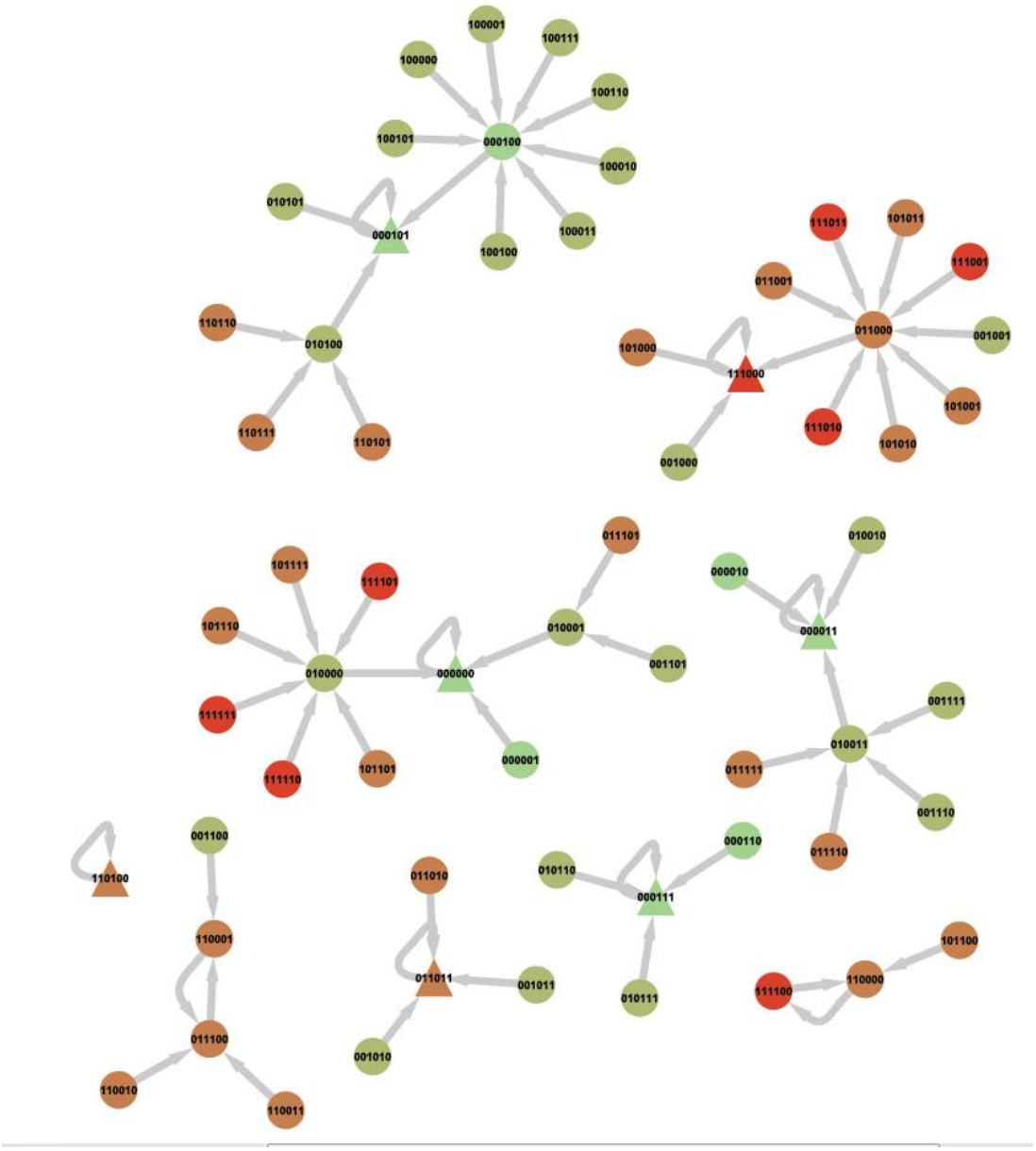


Figure 3.2.: The whole state space of the network by Chickarmane et al. 2008 [68]. The nodes represent the states of the model. Every state is characterized by the values of the system's variables The variables are arranged in the order "OCT4, SOX2, NANOG, GATA6, CDX2, GCNF" and represented by a binary number of 6 digits, every digit representing whether the gene is active (1) or inactive (0). The graphic is shape and color coded. Spherical nodes indicate a transitional state while triangular nodes indicate steady states. Moreover the color stands for the differentiation status of a state. The more a state relates to pluripotency the more it will be red, while it becomes green when approaching a differentiation related state.

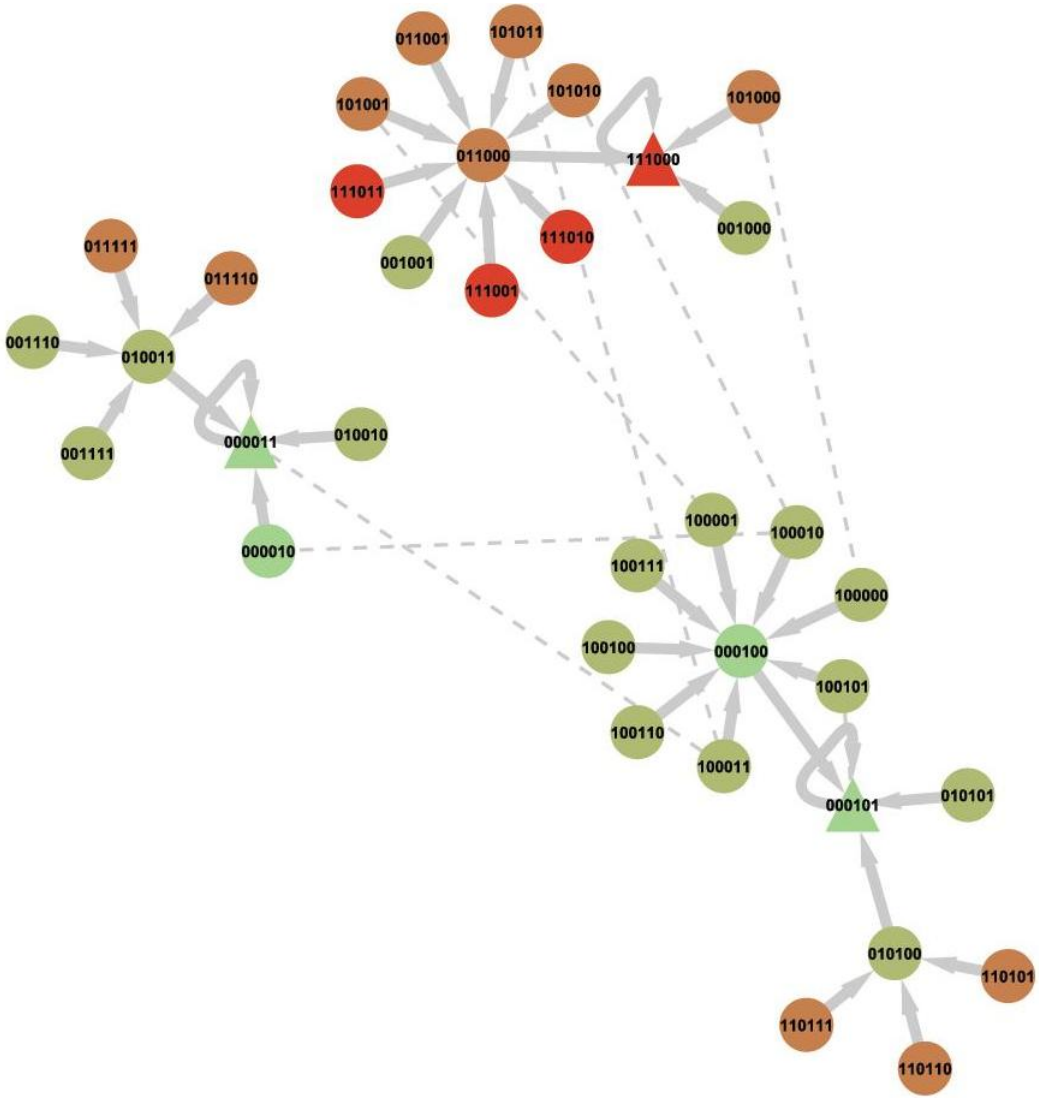


Figure 3.3.: A part of the state space from Figure 3.2 with dashed edges between nodes that have a hamming distance of exactly 1. For simplification and practicability, we only show hamming distances of 1 between nodes that are found in different subnetworks, i.e. that would reach different steady states in time.

# 4. Results

## 4.1. The core network of pluripotency

### The Boolean approach and its limits

We implemented the core GRN of pluripotency as it was first mentioned by Boyer et al. in 2005 and as it is shown in Figure 2.2 and simplified in Figure 4.1 together with its state space. The Boolean functions can be seen in the appendix A.1. As shown in the Boyer publication, OCT4 and SOX2 together form a dimer that binds to the promoter of all three core genes, activating their transcription. In addition, NANOG acts in the same way, thus activating *POU5F1* and *SOX2* transcription as well as the one of its own gene.

When looking at Figure 4.1, we can discern 2 steady states, 000 where all factors are deactivated corresponding to the differentiated state of the cell and 111 where the core network is completely active corresponding to the pluripotency related state. The order in which the species appear in the state vectors is *OCT4*, *SOX2* and *NANOG*. The transition between the two subnetworks made up by the states leading to one or the other steady state, can occur in both direction either through turning *OCT4* and *SOX2* on or off or through activation or deactivation of *NANOG* respectively. These actions being applied on one of the steady states, the system will transition into one state of the other subnetwork leading to the other steady state. This process can then be seen as a reprogramming or differentiation experiment.

When analyzing the network a little closer by looking at the model equations in A.1, we can easily discover that the three species in this Boolean network in fact all have the same equation, thus being equivalent. Hence one can combine these species in one single species that activates its own transcription, thus giving rise to the same dynamics as the network that we just analyzed. This species can then be seen as an ultimate pluripotency switch that remains deactivated in a differentiated state and once activated remains on through self-activation giving rise to a permanent state of pluripotency.

However, since the dynamics of the network suggest that *NANOG* activation, respectively *POU5F1* and *SOX2* activation, alone are sufficient to lead to a transition from the differentiation to the

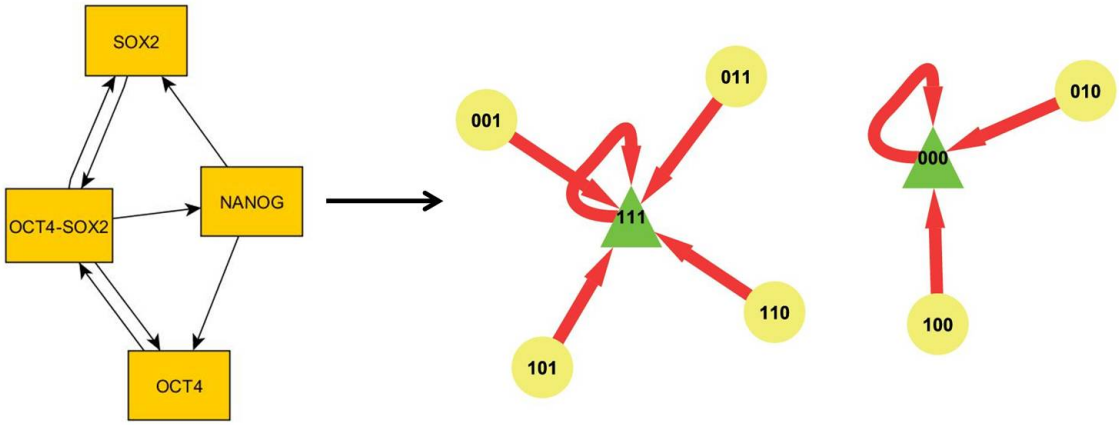


Figure 4.1.: The core network of pluripotency (left) and its state space (right).The species in the state vectors appear in the order OCT4, SOX2, NANOG

pluripotency related steady state, when the biological reality suggests that the delivery of more TFs is necessary to get significant reprogramming efficiencies, it appears that things are more complex. The dynamics of the core transcriptional network alone may be consistent with earlier results [67, 133] showing that OCT4 and SOX2 are necessary for reprogramming, but there are many more mechanisms that are involved in the reprogramming process and that are wired around this core network either upstream or downstream of it or otherwise interconnected with it.

In the following sections we will concentrate on the extension of the core network model, integrating it into different other mechanisms that are known to have a significant effect in iPSC reprogramming.

## 4.2. The epigenetic model, trying to explain the low reprogramming efficiencies or long reprogramming times

### Boolean model

We have mentioned earlier in this work that one of the unsolved problems and questions of iPSC reprogramming consisted in the long times needed for the full process. Simultaneously it is well established that epigenetic pathways are involved in retroviral silencing in iPSCs [81]. Our aim was to include epigenetic modifications into the network of the core transcriptional regulatory

circuitry in order to gain an insight into possible epigenetic mechanisms and implications which could affect the reprogramming time period. Indeed, this question should be approached using ODE modeling after having analyzed the steady states and Boolean dynamics of the network.

Starting from the transcriptional core regulatory network as proposed by Boyer et al. [60] and shown in Figure 2.2, we first included the incorporated reprogramming genes that we named *OCT4*i, *SOX2*i and *NANOG*i into the network since we clearly wanted to separate these two concepts to show which genes are turned on and which are turned off during the reprogramming. Although the encoded transcription factors of course are the same, the endogenous and exogenous genes differ in their promoter and thus in their regulatory input pattern. The second assumption that we made was the incorporation of two genes responsible for epigenetic modifications one of them keeping the endogenous core transcriptional factors inactive in a differentiated state and the other one silencing the retroviral exogenous factors after the reprogramming is done. The one that we called M1 (see Figure 4.2), that controls the expression of the endogenous core regulatory genes, can be identified either with repressive or bivalent histone modifications or as well with DNA methylation at the pluripotency gene promoters, e.g. through Dnmt3a/b, both mechanisms having been mentioned in 2.1.2 and 2.2.2. The species M2 can represent several different mechanisms that are involved in retroviral silencing and that can be looked up in [134].

Moreover, for better visibility and comprehension we combined the species OCT4 and SOX2 to their dimer, as it has been done in Chickarmane et al. 2008 for example [68]. The resulting GRN is shown in Figure 4.2. As this network contains 7 species and many interactions between them, its state space would be made up by 128 nodes, the implementation would be rather complicated and the update rules would probably not perfectly reflect the situation. When taking a closer look at this network, one can see, that *OCT4-SOX2* and *NANOG* on the one hand, as well as *OCT4*i, *SOX2*i and *NANOG*i on the other hand have each the same inputs from the same species as well as the same outputs. As we already eliminated the problem of dimerization of *OCT4* and *SOX2* through eliminating *OCT4* and *SOX2* and only keeping *OCT4-SOX2* as proper species, we can see that the species have the same update rules. Thus, they are equivalent in the network and can be combined into one species that we will call OSN and OSNi respectively. We have analyzed the more complex network as well as the simplified version and found that the simplifications did not change the qualitative behavior of the network and the steady states of the system (results not shown).

This rather easy assumption, even though not respecting the exact dynamics of the biological system, preserves the qualitative behavior of the network and thus leads us to an extremely simplified model with only 4 species and no more than 8 interactions which is shown in Figure 4.3.

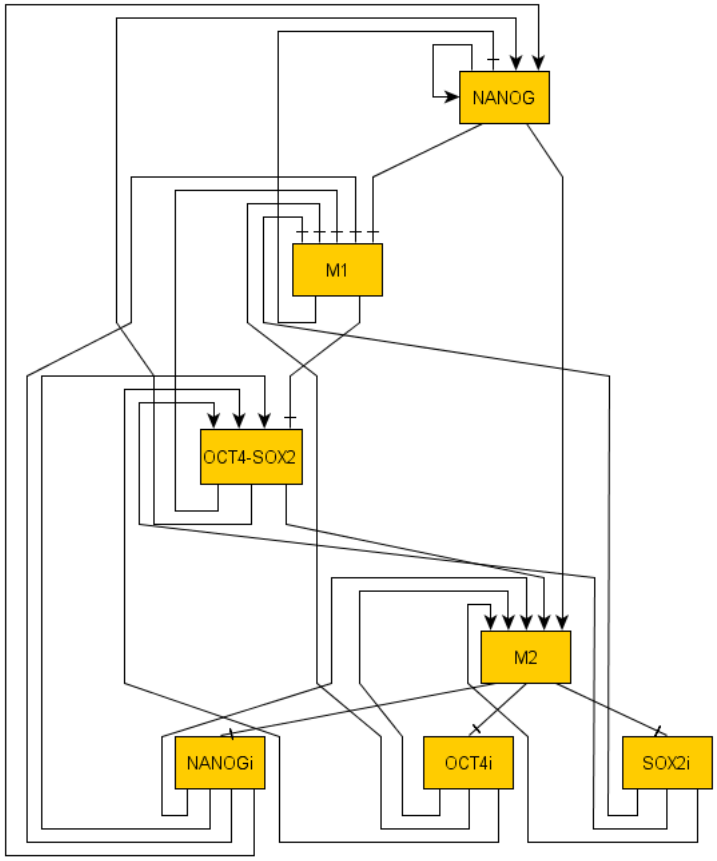


Figure 4.2.: First, large epigenetic model with 7 species that are highly connected through different interactions. Since some of the species can be assumed as equivalent from just looking at their interactions, this model can be strongly simplified which is shown in Figure 4.3

What is still hidden in the two graphics is that in fact, the update rule for OSNi also contains that once it is brought into the cell, it will remain there, transcriptionally active unless M2 is active. So, we included in the update rule, even though not shown in the graphical representation of the network, that OSNi is self sustaining, thus is autoactivating. In fact, if OSNi is present in the cell at time t and there is no inhibition, in our case by M2, then it will stay active in the cell because of its specific promoter. The exact Boolean update rules as the are implemented in BooleanNet are shown in A.2.

Since the model only contains 4 species, there are only $2^4 = 16$ nodes in the state space which is shown in Figure 4.4. In this example, we abandoned the idea of color coding for the state space as it was mentioned in 3.2 because the species can not that easily be classified into categories.

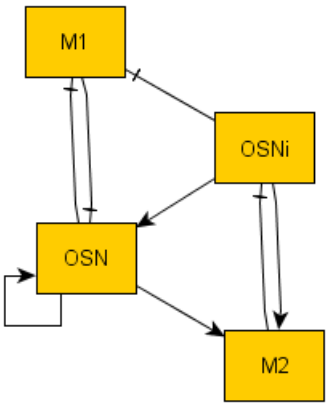


Figure 4.3.: GRN for a simplified epigenetic model. We merged the endogenous core transcriptional factors OCT4, SOX2 and NANOG in one species, OSN and the exogenous retroviral corresponding genes in OSNi. Furthermore, we included the two unspecific epigenetic modifying genes M1 and M2 that both have inhibitory character.

As one can easily see, the system has 2 steady states corresponding to the pluripotency related state where OSN and M2 are activated and OSNi and M1 are silent. In this configuration, the reprogramming is completely finished, the retroviral genes are silenced and the epigenetic modifications represented by M1 that were destined to keep OSN repressed have been turned off. We will first look at the state 0110 in the lower left of Figure 4.4 which corresponds to the introduction of the retroviral genes into a differentiated cell which is in the steady state 0010 since M1 therein maintains the pluripotency genes OSN silenced. This transition that corresponds to a reprogramming experiment is shown by one of the dashed lines in Figure 4.4. We can easily see that this state 0110 will transition in 2 steps into the pluripotency related state 1010 where M1 is turned off, the endogenous pluripotency core genes are activated and the retroviral exogenous genes have been silenced by M2 which is active. This transition from 0010 over 0110 which then ends up in 1001 reproduces in an easy Boolean model including epigenetics the reprogramming process thereby partly justifying our idea. Moreover, the model also predicts that without the exogenous retroviral genes but only by activating the endogenous core transcriptional factors in a transient way like in the state 1010 (Figure 4.4, lower left of the right subnetwork, transition also indicated by a dashed line), the system will not reach a pluripotent state, but fall back into the differentiated steady state 0010.

To show that this epigenetic state space can be remodeled in MS Excel, we implemented the rules for the GRN in the software and tried to reproduce some the paths of the state space. Figure 4.5 shows the implementation and the paths for some initial states but we were able to reconstruct

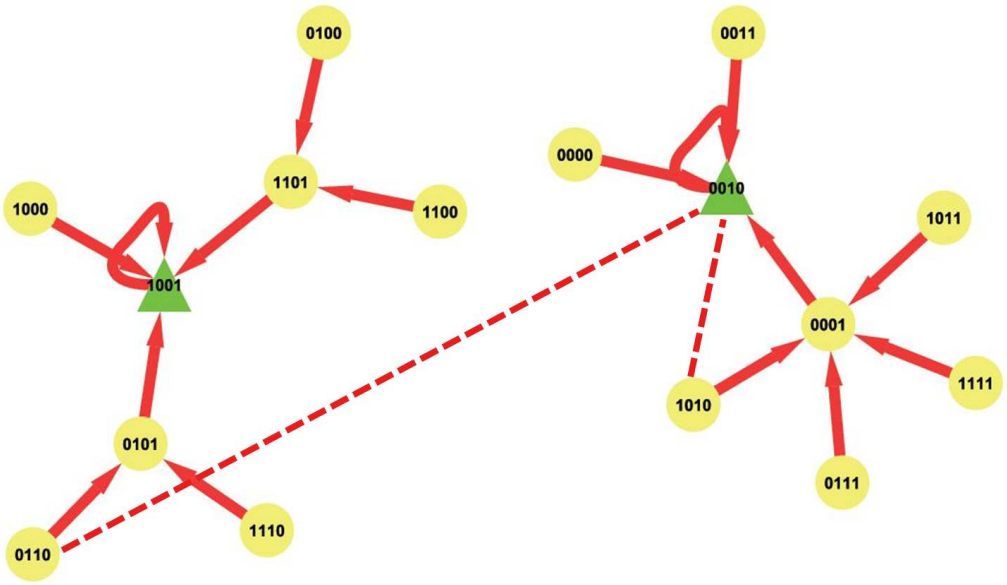


Figure 4.4.: The whole state space of the small epigenetic model from Figure 4.3. The variables of the state vectors are arranged in the order OSN, OSNi, M1 and M2. Thus the two steady states respectively correspond to the pluripotent state (left), where OSN and M2 are active and a differentiated state (right) where only M1 is activated keeping OSN silenced. The dashed lines correspond to significant transitions between nodes with hamming distances of 1 whose implications will be further described in the text.

all the other paths as well (results not shown). As examples, the graphic shows the initial states for the experiment that resembles the reprogramming process as we mentioned above, another one that represents the transient activation of OSN without incorporation of OSNi into the cell which results in a differentiation related state and the last one for the state where all species are active at the beginning. The last path, as mentioned in the graphic as well, shows the importance for the right timing of M2 activation. If it is turned on too early like it is in this special case, it will turn off OSNi which results together with inactivation of OSN in a differentiation related steady state.

Thus, we show that with the use of MS Excel and the combination of BooleanNet and Cytoscape we have two powerful tools designed each for a slightly different purpose. While the implementation of a network structure and the simulation of the time courses for different inputs is very easy and straightforward in MS Excel, the combination of BooleanNet and Cytoscape provides us with a wealth of information and all possible time courses represented in the state space with just a little more effort. However, when the network becomes larger, the size of the state space

increases making it hard to recognize specific paths and information on first sight. For this task, MS Excel seems well suited, because we just have to implement the suited initial conditions to see to which steady state they will lead.

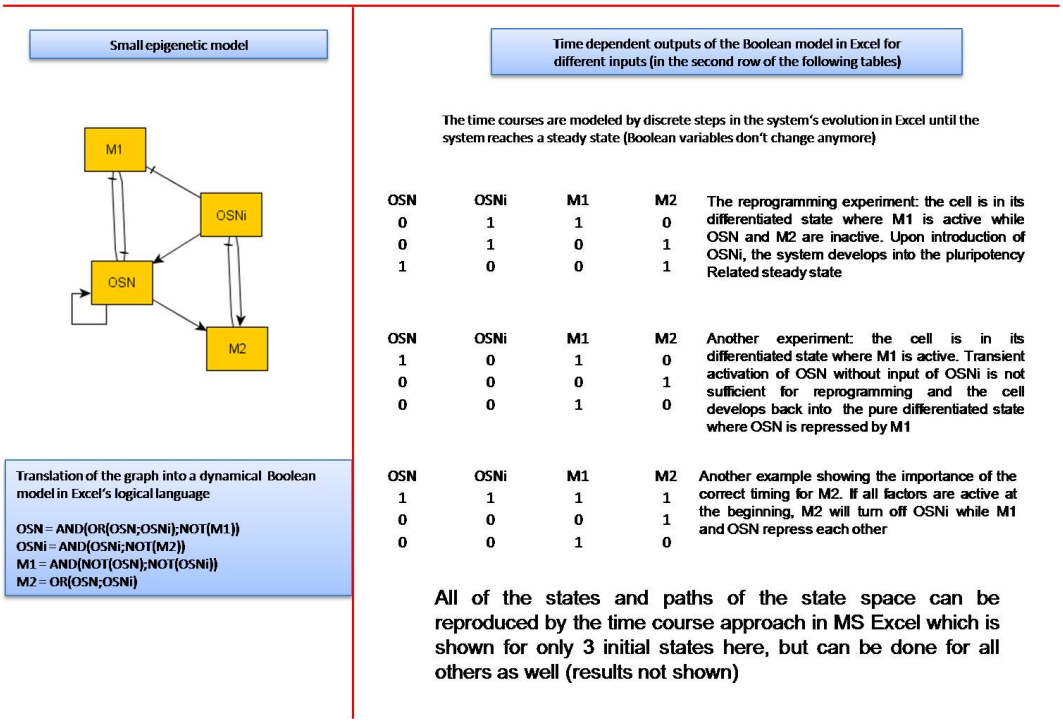


Figure 4.5.: Reproduction of some of the paths of the state space with the MS Excel time course simulation

## Standardized qualitative dynamical model in SQUAD

To show the possibilities, advantages and implications of continuous models compared to discrete models, we implemented the epigenetic model in SQUAD which is described in 3.3.

It is apparent, that when the possibilities to manipulate a network increase, it becomes more complex to work with the software. In the Boolean modeling approach we only had to define the rules that were dictated by the network structure to be able to generate the whole state space. Now, with the use of a software that extrapolates a network structure into a continuous model, we have the possibility to adjust interaction strengths, decay and gain rates and initial conditions for the time course simulation. This is at the same time a big advantage for the analysis of a GRN and a disadvantage for the simplification of network dynamics. The advantage mainly consists in the opportunity given to the analyst by the number of parameters, to manipulate the network

in a way that it fits experimental data and biological reality. However, since now we are dealing with a complex extrapolation method developed by Mendoza and Xenarios in 2006 [123] which has been mentioned in 3.3, it becomes harder for us to retrace the exact mechanisms which lead to the generated outcome, thus making the interpretation of the parameter adjustment difficult.

Nevertheless, we will now present some of the results we got by adjusting the parameters in a specific way and try to interpret them here and below in 5. The parameters are shown in the appendix in tab.A.1.

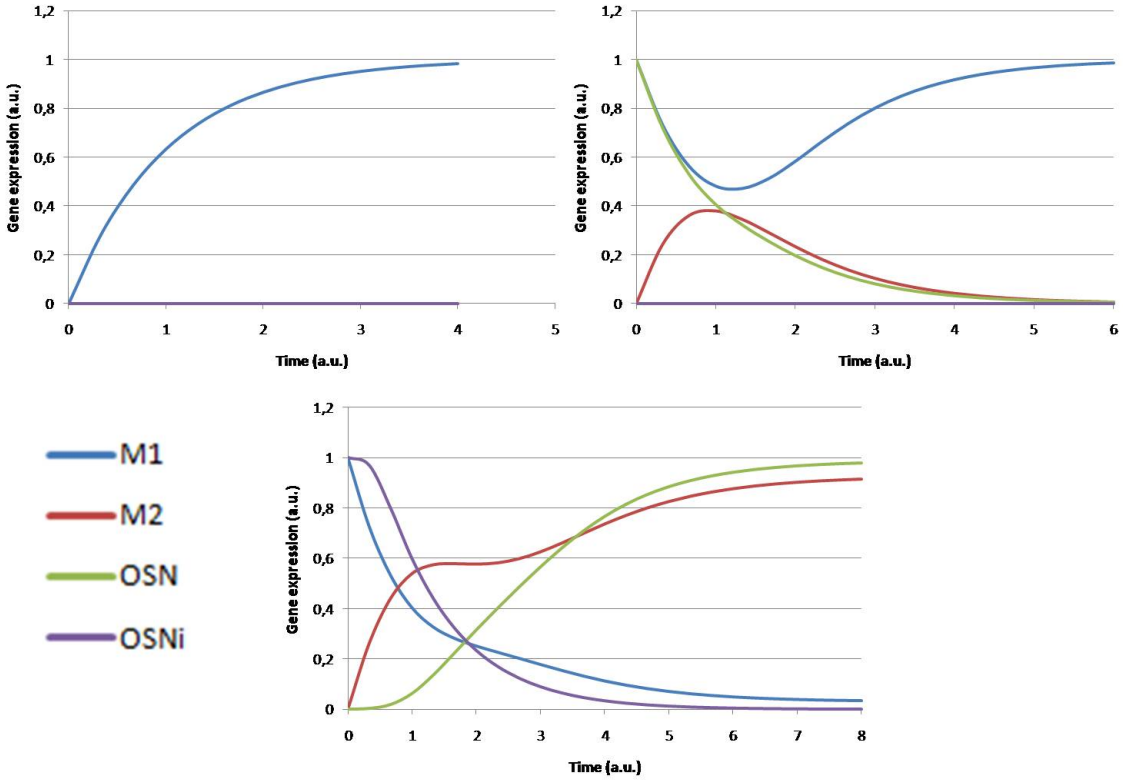


Figure 4.6.: Time courses from SQUAD for different initial conditions of the small epigenetic model. Upper left: When all species are 0 at the beginning, the system reaches the steady state, where M1 is active and OSN, OSNi, M2 are not, the differentiation related state. Upper right: When M1 is active and OSN gets transiently activated with OSNi and M2 inactive, the system reaches the same steady state where only M1 is turned on. Bottom: When M1 and OSNi are active, the system reaches the steady state where OSN and M2 are active, the pluripotency related state.

We generated the time courses, i.e. the graphical representation of the gene's activities over time, for 3 different sets of initial values in SQUAD. Every time course is run until the gene activities reach their steady state. In the upper left of Figure 4.6 we can see the time course when all

genes are deactivated at the beginning corresponding to the state 0000 in the Boolean state space mentioned before. In the same way that 0000 reaches the steady state 0010 corresponding to a differentiated state, the time course simulation as well shows a simple saturation kinetic for M1 reaching its steady state value at about 1, the other species remaining nil. In the upper right right of the same Figure , the system reaches the same steady state when starting from the initial conditions that correspond to the state 1010, i.e. where OSN and M1 are active at the beginning and OSNi and M2 are not. In that case, the dynamical behavior is slightly different. While OSNi stays deactivated, M2 will transiently turn on but decreases afterwards synchronously with OSN while M1 recovers from a transient decrease to its initial steady state value of approximately 1. The last graphic at the bottom of Figure 4.6 finally shows the time course for the initial state corresponding to 0110, i.e. the state where M1 and OSNi are active and which can be interpreted as the initial state for the reprogramming of a differentiated cell. These initial conditions finally lead to the activation of OSN and M2 and the deactivation of M1 and OSNi. As well as the state 0110 in Figure 4.4 reaches the pluripotency related steady state 1001 by activating M2 first and then OSN in the second step when M1 has been downregulated, the dynamic of this initial state in SQUAD shows a similar behavior. In fact, when taking a closer look at the SQUAD simulation, we can see that M1 quickly decreases at the beginning, going over into a slower decrease while the opposite holds for OSNi which decreases in a sigmoidal manner. For our parameter set, this difference is only slightly visible but we can probably increase it with a different parameter set leading to the assumption that maybe OSNi stays active in the beginning for a longer time constantly activating OSN and M2 and inhibiting M1 and and then decreases quickly upon M2 upregulation.

In the following, we will extend our analysis from the standardized qualitative dynamical model to an ODE model where we implement the ODE system on our own.

### ODE model in Copasi

As depicted in 2.3.2 and 3.4 we implemented the epigenetic model from Figure 4.3 in Copasi with the corresponding rate equations leading to a set of 4 ODEs (see A). The conditional definition of the equation for OSNi is due to the fact, that OSNi only underlies an inhibition of the general form

$$\frac{d[OSNi]}{dt} = \frac{V_{max}}{K_m + [M2]^n}$$

thus, it is independent of its own concentration and even if M2 is high, it will be decently expressed. However, if OSNi is not even present in the cell, i.e. its initial concentration is 0, then it won't ever possibly be activated, which is why we had to introduce the condition that if OSNi equals zero, the change of concentration of OSNi vanishes as well.

To reduce the number of variables we neglected the basal transcription rates in a first approach accepting that without activator, the gene will not be expressed at all. Since lack of activation or direct inhibition in our network yields epigenetic modification, this assumption is justified because epigenetically silenced genes are not transcribed anymore. We used the parameters represented in table A.3 in the appendix A.2. Most parameters are just set to 1 except for the Hill coefficients of inhibition which we set to 2 and the degradation rates that were adjusted to 0,1. The higher Hill coefficients for inhibition than for activation reflect somehow the stronger effect of inhibitors that we assumed in the Boolean modeling as well by combining activation and inhibition by an AND gate.

Figure 4.7 shows the resulting time courses for different initial values, either for all factors starting at 0 (upper left), for transient activation of OSN in a differentiated state, i.e. OSN and M1 are high, M2 and OSNi low, or at the beginning of reprogramming with high expression of M1 and OSNi and no expression of M2 and OSN. For the time course in the upper left, where all species are 0 at the beginning, the only species that changes is M1 increasing to its steady state value in a hyperbolic saturation kinetic while the other species remain nil. Thus, we end up in the differentiation related steady state where M1 is expressed and the other species are inactive. The graphic in the upper right shows the time course for high M1 and OSN, and low M2 and OSNi initial concentrations, i.e. for a transient activation of OSN in differentiated cells. While OSNi that lacks these cells stays low during the time course, one can observe, that the concentration of M2 transiently increases through the activation by OSN but decays again upon progressive decrease of OSN. In the opposite biphasic shape, M1 first declines by the action of high OSN, than regenerates after the OSN decay, the system thus ending up in the same differentiation related steady state as before where M1 is highly expressed, while the other factors are kept inactive. Lastly, the bottom time course shows the reprogramming experiment, i.e. high initial values of M1 and OSNi with M2 and OSN being nil. While M2 turns on quickly through the high initial expression of OSNi which is ultimately lowered by M2, M1 decreases allowing OSN to slowly augment. The system ends up in the pluripotency related steady state where OSN and M2 are upregulated and M1 and OSNi are turned off.

The discussed time courses are surprisingly similar to the ones that we got with the implementation of the network structure in SQUAD (see Figure 4.6 just with different time scales due to the arbitrary units of time as we don't have any information on the exact expression timing. In fact,

it is quite astonishing that the results are this similar because as we explained before, SQUAD only uses the network structure to build a qualitative ODE system while in the ODE approach we developed our own ODE system based on a combination of Hill and Michaelis-Menten saturation kinetics. To get similar results by the two approaches is an indicator for the trustworthiness of both the model's implementation and the mathematical methods used.

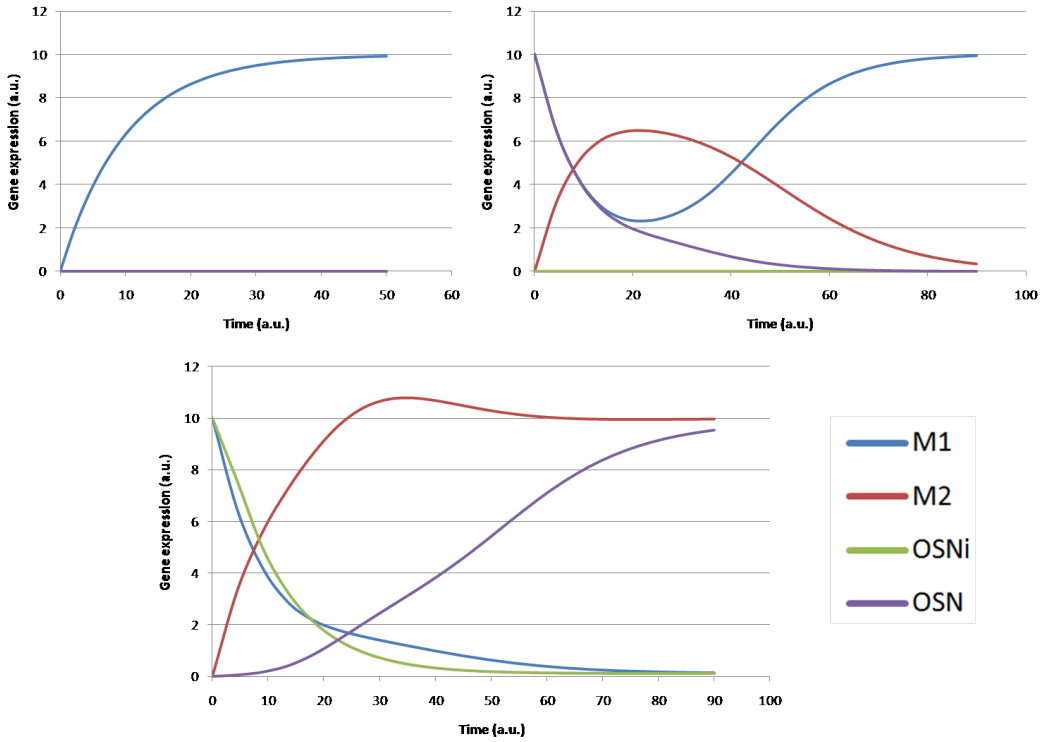


Figure 4.7.: Time courses generated with Copasi for the same initial values as in 4.6. The gene expression and time are in arbitrary units. Upper left: Time course for the initial conditions where all factors are zero at the beginning. Upper right: Initial conditions OSN and M1 are expressed, M2 and OSNi are zero. Bottom: Time course corresponding to the reprogramming experiment where OSNi and M1 are expressed at the beginning and OSN and M2 are not.

## 4.3. The miRNA model, a repressive switch for the pluripotency core network

The idea to construct a network model for the dynamics of pluripotency including miRNAs evolved with recent findings that certain of these small RNAs sequences influence the repro-

gramming efficiency of *Oct4, Sox2 and Klf4 [135]* or directly interact with core transcription factors of the pluripotency network thereby repressing the ESC state [46, 136].

We build a network structure out of the data we got from Alexander Hahn from the Genomatix group. They had evaluated literature and sequence data to establish network structures for each of the four reprogramming genes that Yu et al. used in 2007 to reprogram human somatic cells to iPSCs [6] namely *OCT4*, *SOX2*, *NANOG* and *LIN28*. They then looked for the overlap of these networks, i.e. genes that are present in more than 2 of the singular structures. This overlap is represented in Figure 4.8.

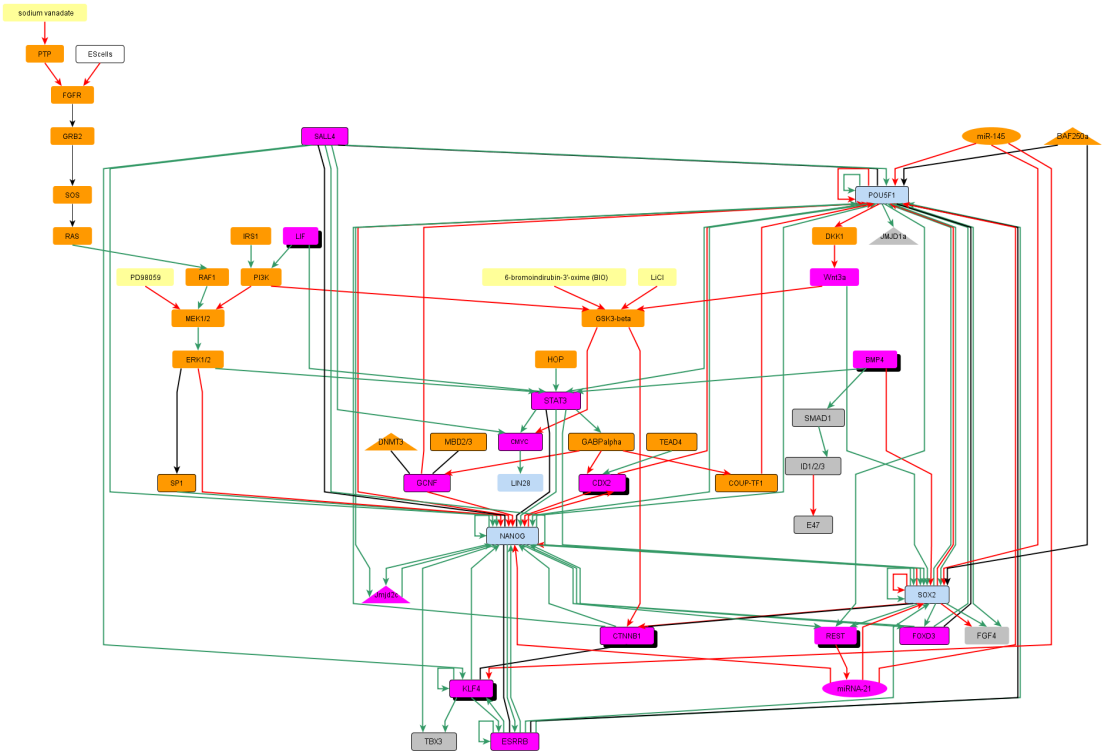


Figure 4.8.: The overlap of genes included in the regulatory networks of OCT4, SOX2, NANOG and LIN28 based upon literature and sequence data as provided by Alexander Hahn from Genomatix.

As we wanted to understand the dynamics of a network responsible for pluripotency, we first eliminated all the external genes that were only inputs to other genes, i.e. weren't affect by any other gene of the network. These genes can be important when considering how to manipulate specific factors of the core network externally but considering the pure dynamics and the analysis of steady states, they do not provide us with further information. Our aim was to build a small network structure including *OCT4*, *SOX2*, *NANOG*, *KLF4* and 2 relevant miRNAs that were

recently found to play an important role in the determination of pluripotency miR-145 [46] and miR-21 [45]. Thus from Figure 4.8 that was manipulated to only contain internal connections between genes we picked only these relevant factors. To get as few species as possible we furthermore eliminated REST from the remaining network. As one can see in Figure 4.9, SOX2, NANOG and OCT4 activate REST that inhibits miR-21 which in turn inhibits the three core factors. For simplification, Figure 4.9 only shows the interactions between the core factors, REST and miR-21 and not the ones between the core factors themselves. Interpreting the graphic to make the network easier, we can assume that OCT4, SOX2 and NANOG directly inhibit miR-21 leaving REST out of the network. As we won't be looking at exact dynamics and exact time-dependent behavior of the network, this assumption is justified because it doesn't matter to us whether REST is upregulated in first place to inhibit miR-21 or if the core factors directly repress the miRNA.

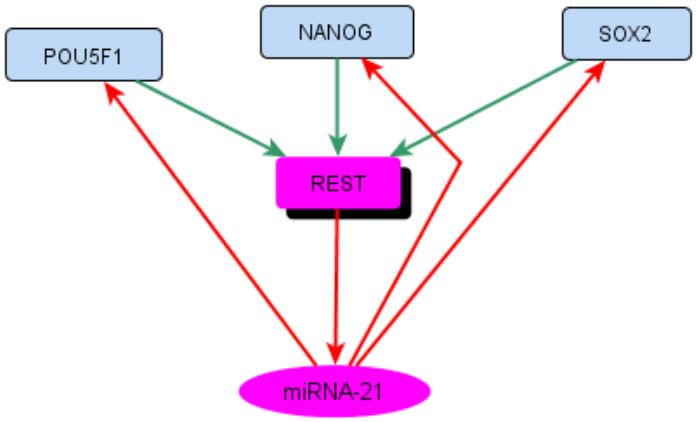


Figure 4.9.: Interactions between OCT4, SOX2, NANOG and miR-21 are mediated by REST which therefore can be left out of the network structure as it is explained in the text

As a next step, to simplify the interactions between the species as far as possible, we eliminated the inhibition of miR-21 by OCT4 and NANOG and let only SOX2 repress miR-21. This may be not trivial to assume but as the network of core factors is supposed to act like a switch, OCT4, SOX2 and NANOG are either all expressed or all inhibited. Furthermore, our Boolean implementation of inhibitions consists in the fact that one inhibition is enough to surpass all the activations and turn a gene off. Understanding these 2 facts, we can thereby admit that in our case, it is not important whether OCT4, SOX2 and NANOG act all three on miR-21 or if it is one of them exclusively. We will demonstrate this effect later in this section when we analyze the robustness of the network by comparing the state spaces of different versions of it.

We then looked for direct inputs to manipulate the network and we found BMP4 and FGF2 to directly act on *SOX2* and *NANOG* respectively. While BMP4 has been found to repress the expression of *SOX2,* FGF2 was shown to stimulate *NANOG* expression levels [137, 111]. Including these interactions into the manipulated network graph yields our final network structure that is shown in Figure 4.10.

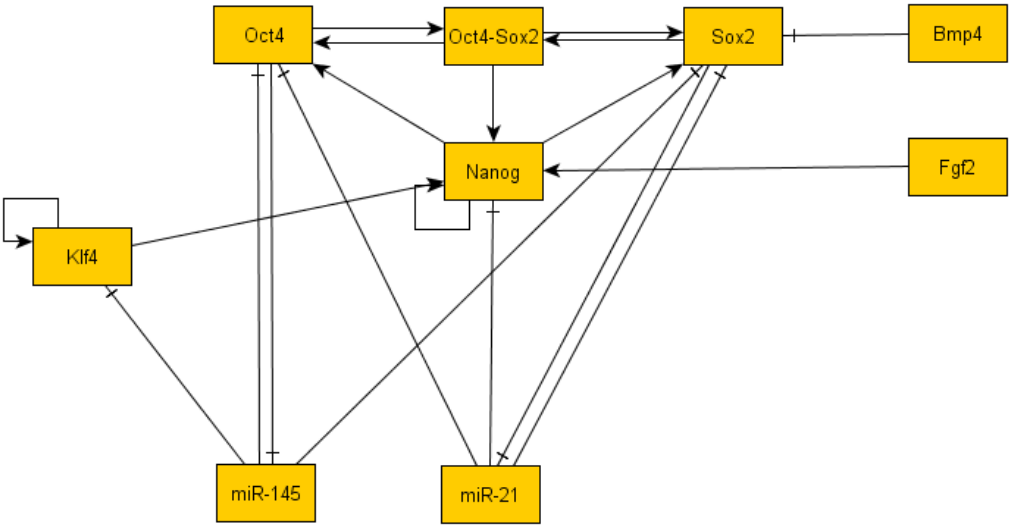


Figure 4.10.: The final GRN structure for the miRNA network including BMP4 and FGF2 as inputs to manipulate the network. Positive interactions are represented by pointed arrows while negative interactions (inhibitions) are indicated by T-shaped arrows.

We first applied the Boolean modeling approach which is introduced in sections 2.3.1 and 3.2. Thus we modeled the state space of the network with BooleanNet and represented it in Cytoscape. This method is explained in detail in the materials and methods section in 3.2. The dimerization of OCT4 and SOX2 for the activation of target genes is not modeled as a proper species but as an AND gate. Instead of creating a new dimer species that would double the number of states in the state space, we just translated the dimerization into the Boolean language by setting the rule for the target gene to: $TG = G1\ AND\ G2$ where TG is the target gene and G1 and G2 are the dimerizing TFs (e.g. OCT4 and SOX2). For more details, see also the materials and methods in chapter 3. The resulting whole state space is shown in Figure 4.11.

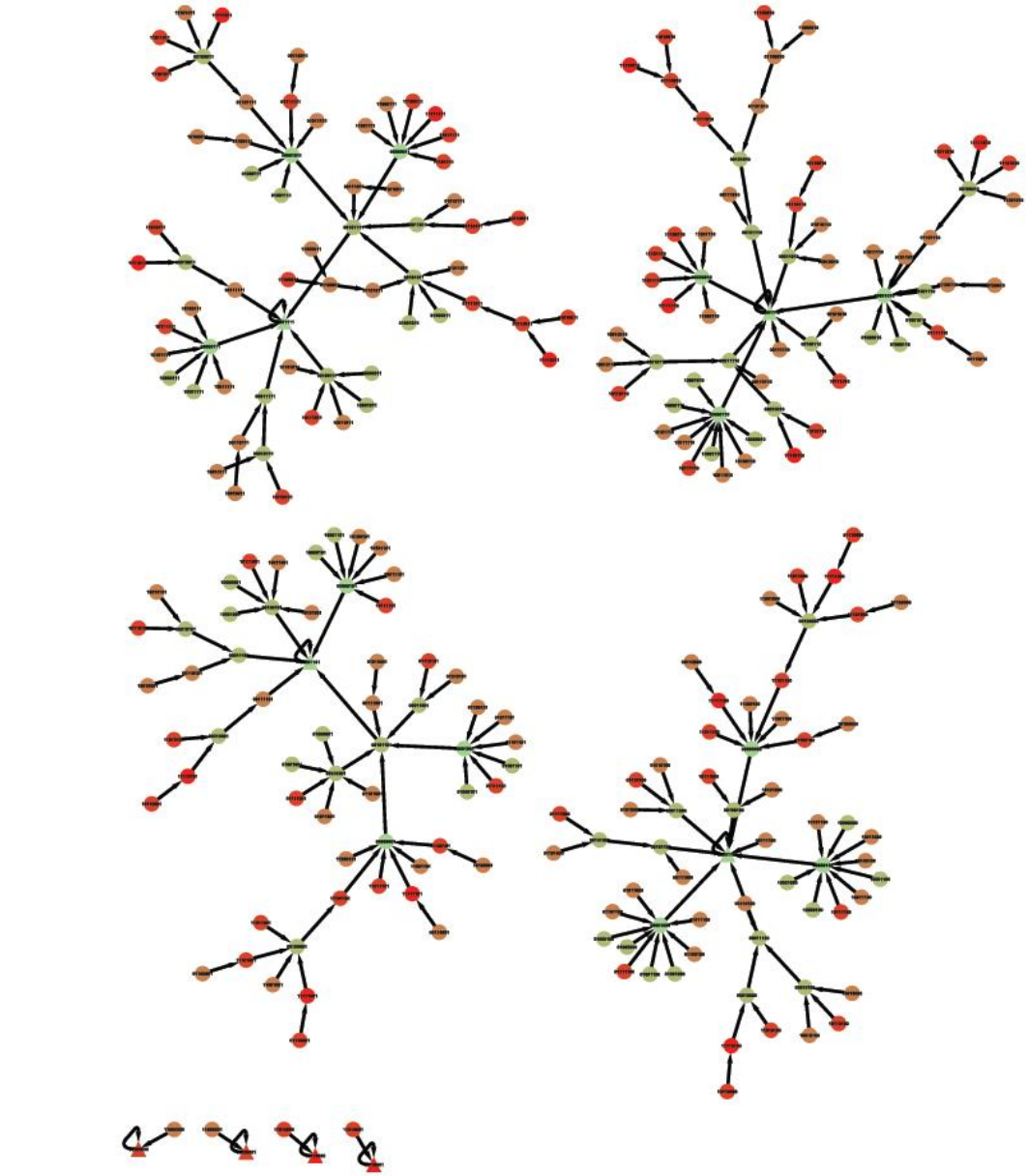

Figure 4.11.: Whole state space of the miRNA network model. States are represented by nodes, transitions are represented by edges between the nodes. A circular node indicates a transitional state, a triangular node indicates a steady state. The different colors of the nodes depend on their pluripotent character. The more pluripotency markers (OCT4, SOX2, NANOG, KLF4) are turned on in a state, the more the color of the nodes turns from green to red and the other way around. We can easily discern 8 clearly separated networks leading to 8 different steady states, 4 of which are related to a differentiated state (the green ones that are attained in the 4 big networks) and 4 that are related to a pluripotent state (the red ones in the left bottom, that are only attained from a few states). For further explanation, see text.

As the network consists of 8 species, there are $2^8 = 256$ states in state space that are indicated by nodes in the state space graph. The edges between the nodes represent transitions between the states. The nodes in the graph have a color and a shape code to make it easier to discern whether the state is related to a differentiated or a pluripotent state in biological reality without having to know the exact value configuration of the state. Hence, circle shaped nodes indicate transitional states while triangular nodes indicate steady states. At the same time, the nodes are color coded depending on the number of pluripotency related genes that are expressed. When all four pluripotency related genes are expressed, the node will be red, if none is expressed, it will be green. The intermediate states take intermediate colors between the two which are generated by the automatic color mapping in Cytoscape.

We can discern at first sight, that the graph consists of 8 subnetworks, 4 of them regrouping all of the states except for 8 states that divide into 4 subnetworks each consisting of 2 states only. The 4 big subnetworks all lead to one of the differentiation related steady states 00001101, 00001100, 00001110 and 00001111. In fact these states have two things in common: in all of them, the pluripotency related genes are repressed while the two miRNAs are expressed. They only differ in the expression of the two input genes *BMP4* and *FGF2*. All combinations of these two input genes lead to 4 different steady states, the same being valid for the whole subnetworks, i.e. all nodes in one subnetwork have the common feature that they have the same expression pattern of the two input genes. In a similar way, the pluripotency related steady states split into the states 11100000, 11100001, 11110000 and 11110001 which have the common features that *OCT4*, *SOX2* and *NANOG* are always active, the miRNA and *BMP4* are repressed and *KLF4* and *FGF2* can switch their expression profile. The miRNAs and *BMP4* acting as pluripotency repressors, it is trivial to say that in a pluripotency related steady state, they have to be inhibited. Moreover, since *KLF4* has an input only from miR-145 which is deactivated in the pluripotent state and from its self-activation, in this situation it can be thought of as an input node to the network just like *FGF2*. Depending on their initial state, they both either remain activated or stay deactivated.

In fact, when leaving *BMP4* and *FGF2* out, the 4 differentiation related subnetworks would merge into one, the same being valid for the pluripotency related networks when we further eliminate *KLF4* we would only get two subnetworks in the whole state space, one leading to a differentiation related steady state, the other one to a pluripotency related one. This can be seen in Figure 4.12 where we represented the state space of the miRNA model reduced by the two inputs BMP4 and FGF2 and by KLF4. We can clearly discern the two steady states 11100 where the core pluripotency genes are active and the miRNAs are not and 00011 where only the miRNAs are active repressing the pluripotency genes. These two steady states span two

subnetworks, one of which, the pluripotency related one, only consists of 2 states, the other one containing the remaining states. Only the pluripotency related steady state itself and the state 11000 corresponding to activation of SOX2 and NANOG directly lead to pluripotency. When we take a closer look at the state 00100 (green node at the top slightly left of the middle axis of Figure 4.12) which corresponds to the activation of NANOG while all other species are inactive, it will transition towards the state 11111, the miRNAs being expressed since NANOG has no repressive control over them. Finally the system will reach the differentiated steady state 00011, because the miRNAs switch off the pluripotency genes. Thus, while in the pure core network, either OCT4 and SOX2 or NANOG had to be expressed, in this case NANOG alone is not sufficient anymore.

Moreover, we represented different transitions between the subnetworks, between nodes that lie at a hamming distance of 1, these transitions accounting for differentiation or reprogramming. There are several transitions that directly lead from the pluripotency related steady state towards states of the differentiation subnetwork, thus representing a disturbance applied on the pluripotent state to trigger differentiation. Activation of anyone of the 2 miRNAs is a possibility for such a transition corresponding to the transitions 11100 –> 11110 and 11100 –> 11101, as well as knockout of either OCT4 or SOX2 as can be seen in the transitions 11100 –> 01100 and 11100 –> 10100. However, there are no hamming transitions that lead from the differentiation related steady state into the pluripotency subnetwork. In fact, we would at least have to input 4 changes into the state 00011 to reach at least 11000 that will then transition into 11100, i.e. both miRNAs would have to be repressed and at least OCT4 and SOX2 activated. Nevertheless, there is the possibility for stochastic reprogramming from one of the transitional states of the subnetwork which will be further discussed in 5.

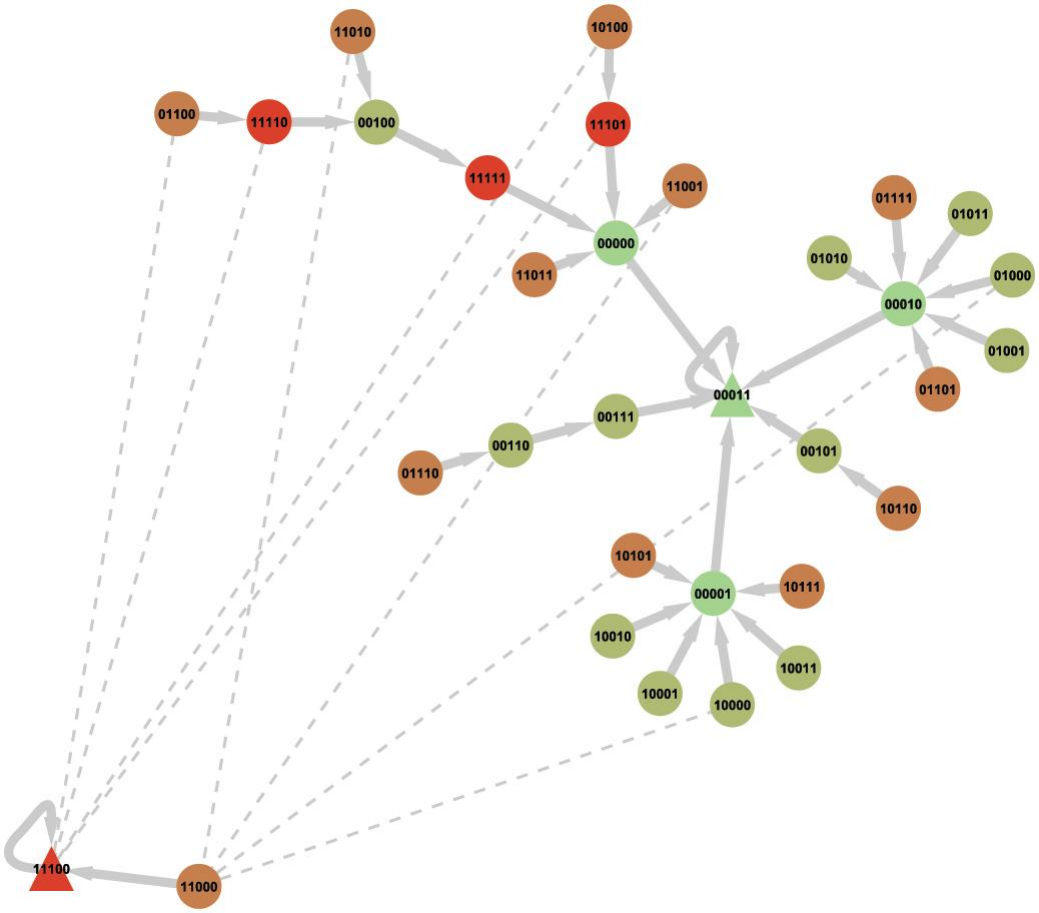


Figure 4.12.: state space with hamming distances for the reduced miRNA model that lacks BMP4, FGF2 and KLF4. The state vectors are arranged in the order OCT4, SOX2, NANOG, miR-145, miR-21. The upper subnetwork leads to the differentiation related steady state 00011, the lower subnetwork consisting of 2 states only leads to the pluripotency related steady state 11100. The dashed lines indicate transitions between states with hamming distance 1.

In fact, when taking a closer look at the network, it is possible to reduce the model even more still preserving the same qualitative behavior. If we proceed in the same way as for the epigenetic model and merge the pluripotency genes in one species as well as the two miRNAs on top of leaving out the input species including KLF4, we will end up in a model with only 2 species mutually inhibiting each other with one of them possessing a self-activation. This model yields a state space with the steady states 10 and 01, the former corresponding to the pluripotency related state which can only be reached by itself, the ladder representing the differentiation related steady states which attracts both other states of the network, i.e. where both species are active or

inactive at the same time. The network and the state space can be seen in Figure 4.13. This state space clearly illustrates the switch like behavior of the miRNAs. In fact, switching the miRNAs on in the pluripotency related steady state, leads to the state 11 which ultimately converts into the differentiation related steady state, while switching off the miRNAs in the differentiation related steady state does not directly lead to the pluripotent state but results in reactivation of the miRNA. To reach the pluripotency state 10 from the differentiation related steady state 01, one either has to repress the miRNA and at the same time deliver the reprogramming genes OSN or to act on one of the states of the bigger subnetwork that is not a steady state either deactivating the miRNAs in the state 11 or delivering the reprogramming genes in 00. However, since most of the cells in a cell population statistically are located in the steady state, this would partly explain the low reprogramming efficiencies because the transition probability is higher for one of the statistically less likely states, i.e. states that are no steady states.

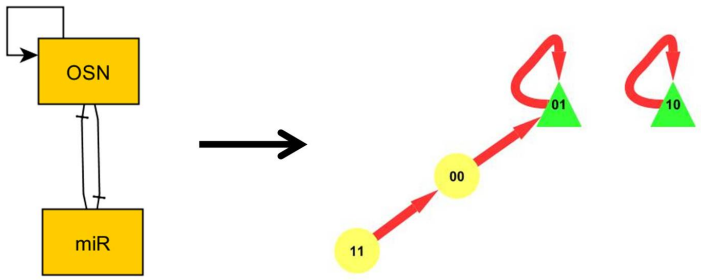


Figure 4.13.: state space (right) for the simple mutual inhibition miRNA network (left)

## 4.4. Including pathways into the pluripotency core network

We have outlined the role of signaling pathways in ESC pluripotency in 2.2.3. Since manipulation of ESCs and generation of iPSCs has become very attractive for possible pharmaceutical use lately and small molecules that act on signaling pathways are abundant, it will be very interesting to know the exact implications, interactions and dynamics of these pathways in relation to the pluripotency core network. More or less recently published results [106, 111, 65, 138] will be used here to create a hypothetical GRN that relates the two branches of TGFß pathway and the FGF pathways and their interaction with the core regulatory master genes of pluripotency. It appears that these pathways play essential roles in the determination of pluripotency and self renewal as depicted in 2.2.3.

We combined again the pluripotency core network as shown in 4.1 to one species that is autoactivating and which we call OSN. The GRN and its Boolean state space are shown in Figure

4.14. In addition to the state space, we also represented a selection of transitions that have a hamming distance of 1 keeping only 8 representatives from the originally 192 transitions for better visibility and understanding.

When looking at the network, we can discern 5 different subnetworks, separated from each other, but only 2 steady states, 000011 and 111100, the former representing the differentiation related steady state where BMP4 and SMAD1/5/8 are active while the other factors are repressed, the ladder representing the pluripotency related steady state where the core regulatory genes of pluripotency as well as TGFß, SMAD2/3 and FGF2 are expressed while BMP4 and SMAD1/5/8 are not. The remaining 4 subnetworks don't end up in a single steady state but in an oscillating attractor, i.e. cycles in the state space graph, which are made up for 3 subnetworks of 4 states and for 1 subnetwork of 2 states. This last subnetwork does not have any transitions with a hamming distance of only 1 to one of the subnetworks containing a steady state and is thus not of great interest for the reprogramming process.

However, by the color code, red for states where more pluripotency genes are turned on, green for states where they are turned off, we can recognize that the subnetworks containing cycles and especially the states of the attractors are more or less related to one or the other tendency. As an example, the subnetwork on the right of Figure 4.14 falls into an attractor that is made up by the state transitions 000111 –> 010001 –> 001010 –> 100011 and back to the state 000111. These states have in common that at most only one factor that can be related to pluripotency is expressed at a time while there is at least 1 inhibitor of pluripotency expressed. Thus, all the 4 states of this attractor are nearer to the differentiated steady state than to the pluripotent one, which can be confirmed by the fact that two of the attractor's states 100011 and 000111 can easily transition directly into the differentiation related steady state which only lies at a hamming distance of 1 from both of the states. In fact, when knocking out OSN in the first state or FGF2 in the second state, we can easily transition into the differentiation related steady state. Another interesting transition from the subnetwork where the attractor is made up by differentiation related states, is the one from 000100 to 100100 which is represented in Figure 4.14 as well in the subnetwork on the left. In fact, the state 100100 is part of the subnetwork containing the pluripotency related steady state 111100 and thus will lead towards it.

When we now focus on the subnetwork at the bottom in the middle of Figure 4.14, we can easily see, that the attractor is made up by the states 110101 –> 011100 –> 111000 –> 101110 that are mainly related to pluripotency. Hence, it is not astonishing that there are states that can directly transition to the pluripotency related steady state (although only one transition is indicated in the Figure, there are many more that are just neglected for better comprehensibility). Nevertheless, what may be a little more surprising is that the subnetwork also contains a state 100000 that

is only at a hamming distance of 1 from the state 000000 contained in the subnetwork leading to the differentiation steady state. However, we have to consider that the species that has to be knocked out here is OSN which in reality represents the three reprogramming factors OCT4, SOX2 and NANOG which would have to be downregulated each of them. This result would have been expected from our analysis of the core network in 4.1 already.

The last subnetwork on the left hand side of Figure 4.14 can be seen as an intermediate between the two subnetworks discussed previously which can also be seen by its mainly brown color that is intermediate between red and green. In fact, the states of the cycle have no clear affinity for one or the other tendencies, i.e. differentiation or pluripotency. Transitions are visible to both subnetworks containing the steady states of the network.

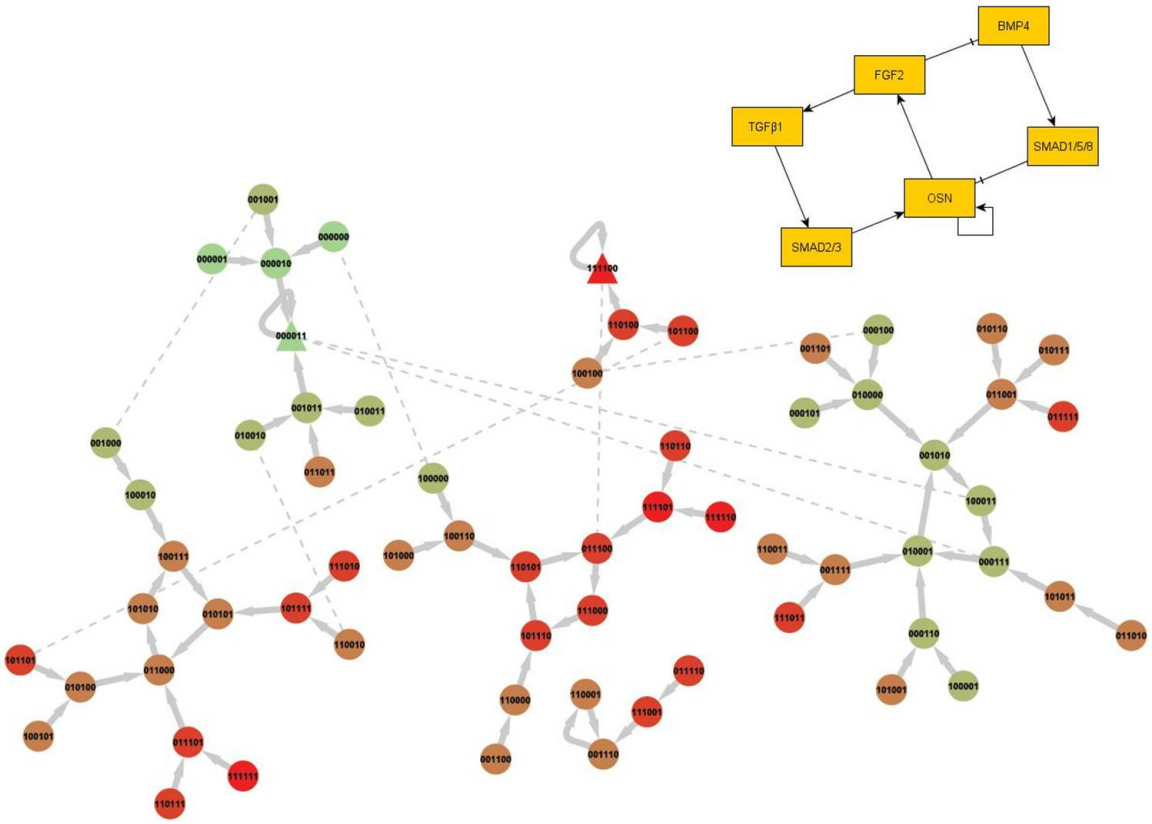


Figure 4.14.: Network structure (upper right) and state space (bottom) with some well chosen transitions between nodes with hamming distance 1. The order with which the species appear in the state vectors is OSN, TGFß1, SMAD2/3, FGF2, BMP4, SMAD1/5/8. The first 4 factors are thought of and color coded to be activators of pluripotency (red), the last two are inhibitors of pluripotency and induce differentiation. The two subnetworks on top of the graphic include the two steady states of the system, while the 4 subnetworks that are below all include cycles, i.e. oscillating attractors.

# 5. Discussion

This thesis constitutes a bilateral work in which on the one hand we presented mathematical approaches and applicable software to model gene regulatory networks and on the other hand analyzed different network structures related to pluripotency and the process of iPSC reprogramming using these methods. Since with ESCs and iPSC reprogramming we were dealing with a very recent research field we decided to keep the mathematical modeling and the software very simple and comprehensible. In fact, experimental data that are required for detailed model construction, ODE modeling and parameter fitting are scarce which is why we approached the problem with a qualitative rather than a quantitative claim. We presented and mathematically analyzed 4 networks in addition to the Chickarmane network that has already been treated in the materials and methods section. In the following we are going to recapitulate the main ideas of the methods for the mathematical modeling and the models that we analyzed.

In this work, we mainly focused on the Boolean modeling of GRNs and especially on the whole state space modeling since we were aiming at finding out about the qualitative behavior of the Boolean networks, i.e. their steady states, cyclic attractors, paths through the network which determine its dynamics and possible transitions between subnetworks that could be related to ESC differentiation or iPSC reprogramming. Thus, we developed a python script for BooleanNet that was able to generate the state space of the network as a directed graph that could be represented and manipulated in Cytoscape. We furthermore introduced the concept of hamming distances for transitions between nodes in the state space that only differ in one digit, i.e. in the expression value of 1 species. In combination with dynamical Boolean modeling using MS Excel as the simplest and most straightforward tool to model Boolean networks, we were able to understand the qualitative behavior of several different networks and generate dynamics that could be reproduced in different continuous models for one of the networks. These continuous models were implemented in a standardized qualitative dynamical model using SQUAD or in an ODE model using Copasi. Although these models are still of qualitative nature because experimental data for the fitting of parameters to get significant results that are consistent with biological reality, are scarce, we were able to generate some of the known qualitative dynamical behaviors as will be discussed in the following.

The first model that we analyzed was the core network of pluripotency as presented by Boyer et al. in 2005 [60] that includes the 3 core genes *OCT4*, *SOX2* and *NANOG* that together span an interconnected self-activation network. We showed this network to have 2 steady states, one being related to a differentiated cell where the core factors are inactive, the other one to a pluripotent cell where the core factors are all permanently expressed. Transitions between these two states were possible either through simultaneous activation of *POU5F1* (expressing OCT4) and *SOX2* or single activation of *NANOG* expression leading to a state that will transition into the pluripotency related steady state. Somehow, NANOG has a stronger effect in this network, it acts as a gatekeeper of pluripotency while OCT4 and SOX2 on their own are not able to turn the network on and thus allow the cell to become an iPSC. Overall the pluripotency core network has the important property to keep its own activity once that it is activated thus allowing the cell to stay pluripotent and self-renew indefinitely in culture.

As mentioned in 2.2.1, the core transcription factors were recently shown to be sensitive to small disturbances and to act in a biphasic manner. In fact, while decent expression of the core factors leads to a pluripotent state, overexpression of them was shown to induce differentiation. Reconstructing this property in Boolean models nevertheless seems impossible because it would require the species to take 3 possible values corresponding to the 3 expression levels, i.e. deactivated, decently expressed and overexpressed, while the Boolean approach only accounts for 2 values per species. Moreover, to generate more complex dynamics that could reflect these properties, we would have to integrate negative feedback that depend on the expression level of the core factors. This has e.g. been done in Chickarmane and Peterson 2008 [68] in an ODE modeling approach, where the OCT4 TF forms a dimer with CDX2 that acts as a repressor on the *POU5F1* and *CDX2* expression. At the same time OCT4 also forms a dimer with GATA6 together repressing *NANOG* expression. Since dimerization is always a concentration dependent process, they could generate a biphasic response of *NANOG* expression as a function of the OCT4-SOX2 dimer concentration, thus partly explaining the repressive effect of their overexpression.

The second model that we have created puts into operation the idea of epigenetic silencing, on the one hand of the reprogramming genes introduced into the cell, on the other hand of the endogenous pluripotency related genes when the cell is in a differentiated state. We have combined the endogenous core regulatory network and the reprogramming genes each into one species and included 2 epigenetic modifiers, acting on each one of them. There is experimental evidence that factors exist which fulfill similar tasks as our hypothetical modifiers, e.g. the DNA methylases Dnmt3a/b or CGBP as described in 2.2.2. We ended up with a simplified model containing only 4 species and 8 interactions shown in. We could show that the network

possesses two steady states, one of which could be identified with fully reprogrammed iPSCs where the endogenous pluripotency related genes maintain their activity, while the exogenous genes are deactivated through epigenetic modifications. The other steady state corresponds to a differentiated cell where the endogenous pluripotency genes are kept repressed by epigenetic modifications. We could show that introduction of the reprogramming genes into the cell could account for reprogramming of the differentiated into the pluripotency related steady state while transient activation of the endogenous core network leads back to the differentiated steady state which is one thing that is known to be true in the biological reality. To be able to reproduce these dynamics with a rather simple model that involves epigenetic modifications which have lately been hypothesized to be responsible for several questions and problems in iPSC reprogramming, is a good indicator for the quality of the model.

Moreover we were able to reconstruct the static and dynamical behavior of the Boolean network in a standardized qualitative dynamical and an ODE model, both of them generating very similar time courses for the initial conditions that we are discussing here, i.e. the reprogramming experiment and the transitional activation of OSN leading back to a differentiated state. It was rather surprising for us that we were able to unify all of these different methods and get similar results for all of them, allowing us to be rather assured of the quality of both the methods used and the implementation of the model.

In addition, the model can also be interpreted in relation with the recently published results of Hanna et al. in December 2009 [14] in the group of Prof. Jaenisch which state that all cells of a cell population can be reprogrammed by the delivery of the reprogramming genes and that this is a stochastic process. In fact, in our Boolean model, every cell will end up in one of the two steady states, the pluripotency related or the differentiation related one. However, upon constant delivery of the reprogramming genes, the differentiation related steady state and thereby all the states that finally lead towards it will end up in the pluripotency related steady state with different latency times, i.e. in a different number of steps, depending on their initial state.

All of these results show that the epigenetic model despite its simplicity may have a great potential to explain different problems and questions of iPSC reprogramming.

The third model included miRNAs that have been mentioned several times lately in the context of ESC research as well. We started from a big network of factors that were found in literature and sequence data to be closely related to the core network of pluripotency. We then step by step simplified the model, reducing the number of species and interactions to finally end up with a network that we started to apply our methods to. The first model revealed us the inconvenience of input nodes in Boolean networks. In fact, the number of states in the state space, as well as the number of subnetworks and steady states increases without providing us with any more

information. As long as the input nodes don't interact, it is thus better to simplify the model until one got rid of every input node to only analyze the behavior of the interconnected GRN. After having done that, we can still choose to include inputs to see which parts of the internal network can be reached from the outside and with which consequences. Reducing the number of species led us to a network with only 2 steady states that could be identified with differentiated or iPS cells respectively. The two resulting subnetworks are very different in size, the differentiation related network regrouping all the transitional states except for two. Transitions through activation or inhibition of one species only lead from the pluripotency related steady state to transitional states of the other subnetwork, but there are no transitions from the differentiation related steady state into the pluripotency related subnetwork. Nevertheless, since transitions from states of the differentiation related subnetwork to the pluripotency subnetwork are possible, there is still the chance for stochastic reprogramming which will be extended further in the discussion below. In fact, while the majority of the cells of course are in one of the two steady states, due to mechanisms such as transcriptional noise, there are still low frequencies of cells in the other states following a Gaussian distribution. Hence, the reprogramming out of these states is possible but the efficiency is very low as will be explained below.

Moreover, one of the biggest qualitative changes compared to the mathematical analysis of the core network from Boyer consists in the fact that the transcription factor NANOG alone is not able anymore to trigger pluripotency when all the other species are deactivated. In fact, this is due to the property that NANOG does not exert any negative control over the miRNAs in contrast to OCT4 and SOX2. When combining the 3 core genes into 1 proper species and the miRNAs as well, we end up in a system with only two species mutually inhibiting each other and one of them having a self-activation loop. Thus, the repressive switch like behavior of the miRNAs becomes even clearer.

The last model dealt with signaling pathways included into the core regulatory network since pathways, especially the two branches of TGFß and FGF have often been attributed a role in reprogramming and differentiation as depicted in 2.2.3. The state space of the model contains 2 steady states one of which can clearly be associated with pluripotency, where the core regulatory genes are activated as well as TGFß, SMAD2/3 and FGF2 and the other one with differentiation where only BMP4 and SMAD1/5/8 are active. Moreover the state space includes 4 cyclic attractors, 2 of which have an intermediate state and 2 others where the cyclic states can again be related to pluripotency and differentiation respectively. We are dealing with a more complex behavior of the system than in the other networks but this is not astonishing, since the pathways are involved in other cellular processes and don't behave in a simple switch like behavior, i.e. they are not exclusively designed for the purpose of pluripotency and self-renewal. Transitions

are mainly possible between similar subnetworks and only few transitional states can cross this barrier. Reprogramming in this context seems difficult and only feasible with simultaneous activation of FGF2 and complete repression or knockout of SMAD1/5/8 which won't be as easy to do because the ladder is involved in many other cellular processes. We haven't found a direct inhibitor of SMAD1/5/8 which in combination with FGF2 would probably be effective for the transition from the differentiated state into the pluripotency related subnetwork. However, the timing seems to be important in this approach as well.

Based on the discussion of our results, we will now try to outline our view of the reasons for the low reprogramming efficiencies or the long reprogramming times. In fact, which has become clear during this discussion is that in our models, the way in the sense of the number of steps for reprogramming of the differentiation related steady state into a pluripotency related steady state is often too long, i.e. there are too many single steps that would have to be taken. Moreover, we have to remember, that even if it was possible, it would be limited to the one model, i.e. epigenetics or miRNAs or pathways, but biological reality is always an interplay of a lot of different mechanisms. Thus, it becomes clear that reprogramming is a long and complicated course of events that has to climb up the energetic hill of all of these features of the different models. In this view of the process, the differentiated state and the pluripotent state are the two states that have a maximal distance from each other. They somehow are the two extremes in which the majority of a population of cells is situated.

However, when introducing the concepts of stochasticity and transcriptional noise which cannot easily be done by Boolean modeling, we can describe a cell population to be in all possible states of a network, each of them having a certain probability and frequency depending on its distance to one of the two steady states. Each of these states has a higher transition probability to the steady state which it is further from than the steady state that it is nearer to itself. Hence, when we look at the states that are near the differentiation related steady state, these higher transition probability states are somehow more poised for reprogramming when compared to the steady state. However, the frequency of cells in one of these intermediate states is very low, because the majority of cells are of course in one of the two steady states. This yields that overall the reprogramming efficiency has to be low, because one of the two things always happens: either a cell is in the differentiation related steady state where reprogramming is nearly impossible because of the many steps that would have to be taken, or a cell is in one of the intermediate states with higher probability, but since there are only very few cells in these states, the overall efficiency is low again.

When trying to increase the reprogramming efficiency it seems necessary that we would either have to enhance all mechanisms that account for pluripotency while repressing all of the ones

that inhibit pluripotency or we would have to increase the number of cells in intermediate states which could be achieved by an increase in the transcriptional noise.

Apart from all the results and features that we could extract from our models, there was one different thing that seemed to be omnipresent as a guideline in our interpretation of the networks. Whenever we generated results of a network, we soon were able to reduce the number of species, thus to simplify the network structure and still preserve the steady states and the rough dynamic behavior of the system. In our opinion this is due to the high connectivity of the pluripotency core network where often species can be thought of as being equivalent which is shown by a similar implementation in the Boolean model. This is another justification for the Boolean modeling approach before transitioning to an ODE model because we would want the ladder to be as simplified and relieved of parameters as possible.

Although we were able to reconstruct the static and dynamical behaviors of the other Boolean networks in ODE systems, we only represented that approach for the epigenetic model because it appears to us as the model with the most potential while being astonishingly simple. Moreover we thought that more detailed analysis by ODE models for every network version would lead to far and go beyond the scope of this work. Thus we mainly focused on Boolean state spaces that present in an easy and comprehensible way many important information on the network.

# 6. Outlook

This thesis gave a qualitative overview of several GRNs including different mechanisms of eukaryotic gene control that have been related to iPSC reprogramming and ESC differentiation. From the beginning, it was thought to provide a broad overview of possibilities to model mechanisms as different as epigenetic modifications and miRNA gene control while still keeping things as simple as possible because the information in this recent research field are scarce and many things are still not well understood avoiding a detailed description of molecular processes by ODEs and parameter fitting to experimental data.

Nevertheless, in addition to the Boolean state spaces, we created standardized qualitative dynamical models and ODE systems to describe the qualitative behavior of the networks although we only showed the results for the small epigenetic network. However, when the data situation will improve in the next years, these models will probably be relevant and useful. We hope that one day our models will be able to be fitted to the parameters and thus making plausible and relevant predictions on the matter of the iPSCs reprogramming timing. In this context, it will be of interest to have time-dependent protein concentration data for the core regulatory genes as well as their epigenetic state during the reprogramming, data on miRNAs and species involved in pathways.

However, until then, we have developed another idea of how to get data for the parameter fitting. We analyzed ChIP-on-chip data together with microarray time courses with the network component analysis (NCA) that is able to calculate the transcription factor strengths of each interaction found by the ChIP-on-chip data [139]. An idea that will possibly be tackled in my future work in the Ph. D. dissertation is to use data that are generated in this way and then somehow normalized, to fit parameters of my models that include genes and transcription factors of the NCA as long as direct protein concentration data are not available.

Another idea that only came a little too late during the writing of this thesis was the extension of the python script for BooleanNet. In fact, it would be interesting what are the shortest paths from the differentiation related steady state to the pluripotency related steady state that only uses a minimal number of hamming transitions, i.e. transitions between nodes with a hamming distance of 1. To continue with the Boolean modeling, I would also like to pursue the stochastic modeling

with the asynchronous method to see to which extent it differs from the simultaneous Boolean modeling and how it could possibly provide us with further information on the networks.

The final aim would be to merge the networks describing the different mechanisms thus creating a holistic systems biology view on pluripotency and reprogramming. However, we have only just begun to understand the basic principles involved in pluripotency and self-renewal, so it is going to be a long but exciting journey.

# A. Model equations

## A.1. Core network model

### Boolean equations

The following rules are update rules, where species with a "*" designate the species at time t+1 and the species without star the ones at time t.

OCT4* = (OCT4 and SOX2) or NANOG

SOX2* = (OCT4 and SOX2) or NANOG

NANOG* = (OCT4 and SOX2) or NANOG

### ODE system

## A.2. Small epigenetic model

### Boolean equations

OSN* = (OSNi or OSN) and not M1

OSNi* = OSNi and not M2

M1* = not (OSN or OSNi)

M2* = OSN or OSNi

### SQUAD parameters

Decay rates are all set to 1,00 while gains are all set to 10,00. The edge weights are represented in tab. A.1.

| Interaction | Weight |
|---|---|
| M1 –\| OSN | 1,00 |
| M2 –\| OSNi | 1,00 |
| OSNi –> M2 | 1,00 |
| OSN –> M2 | 1,00 |
| OSNi –> OSNi | 1,00 |
| OSNi –\| M1 | 2,00 |
| OSN –> OSN | 10,00 |
| OSN –\| M1 | 2,00 |
| OSN –> OSN | 10,00 |

Table A.1.: The weights of the interactions for the epigenetic model in SQUAD

## ODE system

$$\frac{d[M1]}{dt} = V01 + \frac{V11}{(k11 + [OSN]^{n11})(k21 + [OSNi]^{n12})} - k51[M1]$$

$$\frac{d[M2]}{dt} = V02 + \frac{V12 * [OSN]^{n21}}{k12 + [OSN]^{n21}} + \frac{V22 * [OSNi]^{n22}}{k22 + [OSNi]^{n22}} - k52[M2]$$

$$\frac{d[OSNi]}{dt} = \begin{cases} 0, \ M2 = 0 \\ V03 + \frac{V13}{k13 + [M2]^{n31}} - k53[OSNi], \ else \end{cases}$$

$$\frac{d[OSN]}{dt} = V04 + \frac{V14 * [OSN]^{n41}}{(k14 + [OSN]^{n41})(k24 + [M1]^{n42})} + \frac{V24 * [OSNi]^{n43}}{(k34 + [OSNi]^{n43})(k24 + [M1]^{n42})} - k54[M2]$$

| V01 | V02 | V03 | V04 | V11 | V12 | V22 | V13 | V14 | V24 | k11 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| k21 | k12 | k22 | k13 | k14 | k24 | k34 | k51 | k52 | k53 | k54 | n21 | n22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0.1 | 10 | 0.1 | 1 | 1 | 1 | 1 | 1 | 1 |

| n41 | n43 | n11 | n12 | n31 | n42 |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 2 | 2 |

Table A.3.: The parameters of the small epigenetic ODE model used for the time course simulation from Figure 4.7

The Hill coefficients corresponding to activation are all 1, the ones for inhibition all 2, i.e. $n21 = n22 = n41 = n43 = 1$ and $n11 = n12 = n31 = n42 = 2$. For the rest, all degradation parameters are set to 0.1, the basal transcription rates V0i to 0 and the other parameters to 1.

## A.3. miRNA model

### Boolean equations

OCT4* = ((OCT4 and SOX2) or NANOG) and not (miR-145 or miR-21)

SOX2* = ((OCT4 and SOX2) or NANOG) and not (miR-145 or miR-21)

NANOG* = ((OCT4 and SOX2) or NANOG) and not miR-21

KLF4* = KLF4 and not miR-145

miR-145* = not OCT4

miR-21* = not SOX2

BMP4* = BMP4

FGF2* = FGF2

## A.4. The signaling pathway model

OSN* = (OSN or SMAD2) and not SMAD1

TGF* = FGF2

SMAD2* = TGF

FGF2* = OSN

BMP4* = not FGF2

SMAD1* = BMP4

# Eigenständigkeitserklärung

Hiermit versichere ich, Till Scharp, dass ich die vorliegende Diplomarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, den 09.02.2010

# Bibliography


[1] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, Aug 2006.

[2] Jacob Hanna, Marius Wernig, Styliani Markoulaki, Chiao-Wang Sun, Alexander Meissner, John P Cassady, Caroline Beard, Tobias Brambrink, Li-Chen Wu, Tim M Townes, and Rudolf Jaenisch. Treatment of sickle cell anemia mouse model with ips cells generated from autologous skin. *Science*, 318(5858):1920–1923, Dec 2007.

[3] Ning Sun, Nicholas J Panetta, Deepak M Gupta, Kitchener D Wilson, Andrew Lee, Fangjun Jia, Shijun Hu, Athena M Cherry, Robert C Robbins, Michael T Longaker, and Joseph C Wu. Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. *Proc Natl Acad Sci U S A*, 106(37):15720–15725, Sep 2009.

[4] Robert Passier, Linda W van Laake, and Christine L Mummery. Stem-cell-based therapy and lessons from the heart. *Nature*, 453(7193):322–329, May 2008.

[5] R. Dalla-Favera, M. Bregni, J. Erikson, D. Patterson, R. C. Gallo, and C. M. Croce. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in burkitt lymphoma cells. *Proc Natl Acad Sci U S A*, 79(24):7824–7827, Dec 1982.

[6] Junying Yu, Maxim A Vodyanik, Kim Smuga-Otto, Jessica Antosiewicz-Bourget, Jennifer L Frane, Shulan Tian, Jeff Nie, Gudrun A Jonsdottir, Victor Ruotti, Ron Stewart, Igor I Slukvin, and James A Thomson. Induced pluripotent stem cell lines derived from human somatic cells. *Science*, 318(5858):1917–1920, Dec 2007.

[7] Jing Liao, Zhao Wu, Ying Wang, Lu Cheng, Chun Cui, Yuan Gao, Taotao Chen, Lingjun Rao, Siye Chen, Nannan Jia, Huiming Dai, Shunmei Xin, Jiuhong Kang, Gang Pei, and Lei Xiao. Enhanced efficiency of generating induced pluripotent stem (ips) cells from human somatic cells by a combination of six transcription factors. *Cell Res*, 18(5):600–603, May 2008.

[8] Danwei Huangfu, Rene Maehr, Wenjun Guo, Astrid Eijkelenboom, Melinda Snitow, Alice E Chen, and Douglas A Melton. Induction of pluripotent stem cells by defined factors

is greatly improved by small-molecule compounds. *Nat Biotechnol*, 26(7):795–797, Jul 2008.

[9] Keisuke Kaji, Katherine Norrby, Agnieszka Paca, Maria Mileikovsky, Paria Mohseni, and Knut Woltjen. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature*, 458(7239):771–775, Apr 2009.

[10] Keisuke Okita, Masato Nakagawa, Hong Hyenjong, Tomoko Ichisaka, and Shinya Yamanaka. Generation of mouse induced pluripotent stem cells without viral vectors. *Science*, 322(5903):949–953, Nov 2008.

[11] Yan Shi, Caroline Desponts, Jeong Tae Do, Heung Sik Hahm, Hans R Schoeler, and Sheng Ding. Induction of pluripotent stem cells from mouse embryonic fibroblasts by oct4 and klf4 with small-molecule compounds. *Cell Stem Cell*, 3(5):568–574, Nov 2008.

[12] Costas A Lyssiotis, Ruth K Foreman, Judith Staerk, Michael Garcia, Divya Mathur, Styliani Markoulaki, Jacob Hanna, Luke L Lairson, Bradley D Charette, Laure C Bouchez, Michael Bollong, Conrad Kunick, Achim Brinker, Charles Y Cho, Peter G Schultz, and Rudolf Jaenisch. Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of klf4. *Proc Natl Acad Sci U S A*, 106(22):8912–8917, Jun 2009.

[13] Dohoon Kim, Chun-Hyung Kim, Jung-Il Moon, Young-Gie Chung, Mi-Yoon Chang, Baek-Soo Han, Sanghyeok Ko, Eungi Yang, Kwang Yul Cha, Robert Lanza, and Kwang-Soo Kim. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell*.

[14] Jacob Hanna, Krishanu Saha, Bernardo Pando, Jeroen van Zon, Christopher J Lengner, Menno P Creyghton, Alexander van Oudenaarden, and Rudolf Jaenisch. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, 462(7273):595–601, Dec 2009.

[15] A. Ralston and J. Rossant. Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet*, 68(2):106–112, Aug 2005.

[16] J. H. Wanscher. The history of wilhelm johannsen's genetical terms and concepts from the period 1903 to 1926. *Centaurus*, 19(2):125–147, 1975.

[17] Helen Pearson. Genetics: what is a gene? *Nature*, 441(7092):398–401, May 2006.

[18] Kaiser Krieger Scott Bretscher Ploegh Matsudaira Lodish, Berk. *Molecular cell biology*. W H Freeman and Company, 1999.

[19] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, Aug 1970.

[20] F. H. CRICK. On protein synthesis. *Symp Soc Exp Biol*, 12:138–163, 1958.

[21] J.D. Watson. *Molecular biology of the gene*. W.A. Benjamin, 1965.

[22] Knust Janning. *Genetik*. Georg Thieme Verlag, 2004.

[23] Ali H Brivanlou and James E Darnell. Signal transduction and the control of gene expression. *Science*, 295(5556):813–818, Feb 2002.

[24] Craig L Peterson and Marc-Andre Laniel. Histones and histone modifications. *Curr Biol*, 14(14):R546–R551, Jul 2004.

[25] Tony Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693–705, Feb 2007.

[26] K. Luger, A. W. Maeder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648):251–260, Sep 1997.

[27] A. S. Belmont and K. Bruce. Visualization of g1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol*, 127(2):287–302, Oct 1994.

[28] T. M. Fletcher and J. C. Hansen. Core histone tail domains mediate oligonucleosome folding and nucleosomal dna organization through distinct molecular mechanisms. *J Biol Chem*, 270(43):25359–25362, Oct 1995.

[29] P. M. Schwarz, A. Felthauser, T. M. Fletcher, and J. C. Hansen. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry*, 35(13):4009–4015, Apr 1996.

[30] Joanna Wysocka, Tomek Swigut, Hua Xiao, Thomas A Milne, So Yeon Kwon, Joe Landry, Monika Kauer, Alan J Tackett, Brian T Chait, Paul Badenhorst, Carl Wu, and C. David Allis. A phd finger of nurf couples histone h3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442(7098):86–90, Jul 2006.

[31] Xiaobing Shi, Tao Hong, Kay L Walter, Mark Ewalt, Eriko Michishita, Tiffany Hung, Dylan Carney, Pedro Pena, Fei Lan, Mohan R Kaadige, Nicolas Lacoste, Christelle Cayrou, Foteini Davrazou, Anjanabha Saha, Bradley R Cairns, Donald E Ayer, Tatiana G Kutateladze, Yang Shi, Jacques Cote, Katrin F Chua, and Or Gozani. Ing2 phd domain links histone h3 lysine 4 methylation to active gene repression. *Nature*, 442(7098):96–99, Jul 2006.

[32] Ying Huang, Jia Fang, Mark T Bedford, Yi Zhang, and Rui-Ming Xu. Recognition of histone h3 lysine-4 methylation by the double tudor domain of jmjd2a. *Science*, 312(5774):748–751, May 2006.

[33] Marilyn G Pray-Grant, Jeremy A Daniel, David Schieltz, John R Yates, and Patrick A Grant. Chd1 chromodomain links histone h3 methylation with saga- and slik-dependent acetylation. *Nature*, 433(7024):434–438, Jan 2005.

[34] Christopher R Vakoc, Sean A Mandat, Benjamin A Olenchock, and Gerd A Blobel. Histone h3 lysine 9 methylation and hp1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell*, 19(3):381–391, Aug 2005.

[35] Bradley E. Bernstein, Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, Alex Meissner, Marius Wernig, Kathrin Plath, Rudolf Jaenisch, Alex, re Wagschal, Robert Feil, Stuart L. Schreiber, Eric S. L, and er. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326, April 2006.

[36] Veronique Azuara, Pascale Perry, Stephan Sauer, Mikhail Spivakov, Helle F Jorgensen, Rosalind M John, Mina Gouti, Miguel Casanova, Gary Warnes, Matthias Merkenschlager, and Amanda G Fisher. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*, 8(5):532–538, May 2006.

[37] R. C. Lee, R. L. Feinbaum, and V. Ambros. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell*, 75(5):843–854, Dec 1993.

[38] G. Ruvkun. Molecular biology. glimpses of a tiny rna world. *Science*, 294(5543):797–799, Oct 2001.

[39] Eric C Lai. microrna: runts of the genome assert themselves. *Curr Biol*, 13(23):R925–R936, Dec 2003.

[40] David P Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, Jan 2004.

[41] Victor Ambros. Microrna pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, 113(6):673–676, Jun 2003.

[42] P. H. Olsen and V. Ambros. The lin-4 regulatory rna controls developmental timing in caenorhabditis elegans by blocking lin-14 protein synthesis after the initiation of translation. *Dev Biol*, 216(2):671–680, Dec 1999.

[43] V. Narry Kim and Jin-Wu Nam. Genomics of microrna. *Trends Genet*, 22(3):165–173, Mar 2006.

[44] Agnieszka Rybak, Heiko Fuchs, Lena Smirnova, Christine Brandt, Elena E Pohl, Robert Nitsch, and F. Gregory Wulczyn. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol*, 10(8):987–993, Aug 2008.

[45] Sanjay K Singh, Mohamedi N Kagalwala, Jan Parker-Thornburg, Henry Adams, and Sadhan Majumder. Rest maintains self-renewal and pluripotency of embryonic stem cells. *Nature*, 453(7192):223–227, May 2008.

[46] Na Xu, Thales Papagiannakopoulos, Guangjin Pan, James A. Thomson, and Kenneth S. Kosik. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*.

[47] F. JACOB and J. MONOD. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3:318–356, Jun 1961.

[48] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–3297, Dec 1998.

[49] Ian Chambers, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith. Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–655, May 2003.

[50] David C Hay, Linda Sutherland, John Clark, and Tom Burdon. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells*, 22(2):225–235, 2004.

[51] Kaoru Mitsui, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *Cell*, 113(5):631–642, May 2003.

[52] J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, and A. Smith. Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. *Cell*, 95(3):379–391, Oct 1998.

[53] Holm Zaehres, M. William Lensch, Laurence Daheron, Sheila A Stewart, Joseph Itskovitz-Eldor, and George Q Daley. High-efficiency rna interference in human embryonic stem cells. *Stem Cells*, 23(3):299–305, Mar 2005.

[54] Maryam M Matin, James R Walsh, Paul J Gokhale, Jonathan S Draper, Ahmad R Bahrami, Ian Morton, Harry D Moore, and Peter W Andrews. Specific knockdown of oct4 and beta2-microglobulin expression by rna interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells*, 22(5):659–668, 2004.

[55] Louise Hyslop, Miodrag Stojkovic, Lyle Armstrong, Theresia Walter, Petra Stojkovic, Stefan Przyborski, Mary Herbert, Alison Murdoch, Tom Strachan, and Majlinda Lako. Downregulation of nanog induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells*, 23(8):1035–1043, Sep 2005.

[56] Joon-Lin Chew, Yuin-Han Loh, Wensheng Zhang, Xi Chen, Wai-Leong Tam, Leng-Siew Yeap, Pin Li, Yen-Sin Ang, Bing Lim, Paul Robson, and Huck-Hui Ng. Reciprocal transcriptional regulation of pou5f1 and sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell. Biol.*, 25(14):6031–6046, July 2005.

[57] Ariel A Avilion, Silvia K Nicolis, Larysa H Pevny, Lidia Perez, Nigel Vivian, and Robin Lovell-Badge. Multipotent cell lineages in early mouse development depend on sox2 function. *Genes Dev*, 17(1):126–140, Jan 2003.

[58] Helen Fong, Kristi A Hohenstein, and Peter J Donovan. Regulation of self-renewal and pluripotency by sox2 in human embryonic stem cells. *Stem Cells*, 26(8):1931–1938, Aug 2008.

[59] H. Niwa, J. Miyazaki, and A. G. Smith. Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nat Genet*, 24(4):372–376, Apr 2000.

[60] Laurie A Boyer, Tong Ihn Lee, Megan F Cole, Sarah E Johnstone, Stuart S Levine, Jacob P Zucker, Matthew G Guenther, Roshan M Kumar, Heather L Murray, Richard G Jenner, David K Gifford, Douglas A Melton, Rudolf Jaenisch, and Richard A Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956, Sep 2005.

[61] Takao Kuroda, Masako Tada, Hiroshi Kubota, Hironobu Kimura, Shin ya Hatano, Hirofumi Suemori, Norio Nakatsuji, and Takashi Tada. Octamer and sox elements are required for transcriptional cis regulation of nanog gene expression. *Mol Cell Biol*, 25(6):2475–2485, Mar 2005.

[62] David J Rodda, Joon-Lin Chew, Leng-Hiong Lim, Yuin-Han Loh, Bei Wang, Huck-Hui Ng, and Paul Robson. Transcriptional regulation of nanog by oct4 and sox2. *J Biol Chem*, 280(26):24731–24737, Jul 2005.

[63] Guangjin Pan and James A Thomson. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res*, 17(1):42–49, Jan 2007.

[64] Yasmin Babaie, Ralf Herwig, Boris Greber, Thore C. Brink, Wasco Wruck, Detlef Groth, Hans Lehrach, Tom Burdon, and James Adjaye. Analysis of Oct4-Dependent transcriptional networks regulating Self-Renewal and pluripotency in human embryonic stem cells. *Stem Cells*, 25(2):500–510, February 2007.

[65] Xi Chen, Han Xu, Ping Yuan, Fang Fang, Neil D Clarke, Chia-Lin Wei, and Huck-Hui Ng. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–1117, Jun 2008.

[66] Qing Zhou, Hiram Chipperfield, Douglas A Melton, and Wing Hung Wong. A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*, 104(42):16438–16443, Oct 2007.

[67] Natalia Ivanova, Radu Dobrin, Rong Lu, Iulia Kotenko, John Levorse, Christina DeCoste, Xenia Schafer, Yi Lun, and Ihor R Lemischka. Dissecting self-renewal in stem cells with rna interference. *Nature*, 442(7102):533–538, Aug 2006.

[68] Vijay Chickarmane and Carsten Peterson. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS ONE*, 3(10):e3478, 2008. PMC2566811.

[69] Hitoshi Niwa, Yayoi Toyooka, Daisuke Shimosato, Dan Strumpf, Kadue Takahashi, Rika Yagi, and Janet Rossant. Interaction between oct3/4 and cdx2 determines trophectoderm differentiation. *Cell*, 123(5):917–929, December 2005.

[70] Y. I. Yeom, G. Fuhrmann, C. E. Ovitt, A. Brehm, K. Ohbo, M. Gross, K. Hubner, and H. R. Scholer. Germline regulatory element of oct-4 specific for the totipotent cycle of embryonal cells. *Development*, 122(3):881–894, Mar 1996.

[71] Boris Greber, Hans Lehrach, and James Adjaye. Silencing of core transcription factors in human ec cells highlights the importance of autocrine fgf signaling for self-renewal. *BMC Developmental Biology*, 7(1):46, 2007.

[72] Janel L Kopp, Briana D Ormsbee, Michelle Desler, and Angie Rizzino. Small increases in the level of sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells*, 26(4):903–911, Apr 2008.

[73] Brian Boer, Janel Kopp, Sunil Mallanna, Michelle Desler, Harini Chakravarthy, Phillip J Wilder, Cory Bernadt, and Angie Rizzino. 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 Res*, 35(6):1773–1786, 2007.

[74] Masato Nakagawa, Michiyo Koyanagi, Koji Tanabe, Kazutoshi Takahashi, Tomoko Ichisaka, Takashi Aoi, Keisuke Okita, Yuji Mochiduki, Nanako Takizawa, and Shinya Yamanaka. Generation of induced pluripotent stem cells without myc from mouse and human fibroblasts. *Nat Biotechnol*, 26(1):101–106, Jan 2008.

[75] Yuin-Han Loh, Jia-Hui Ng, and Huck-Hui Ng. Molecular framework underlying pluripotency. *Cell Cycle*, 7(7):885–891, Apr 2008.

[76] Yuin-Han Loh, Qiang Wu, Joon-Lin Chew, Vinsensius B Vega, Weiwei Zhang, Xi Chen, Guillaume Bourque, Joshy George, Bernard Leong, Jun Liu, Kee-Yew Wong, Ken W Sung, Charlie W H Lee, Xiao-Dong Zhao, Kuo-Ping Chiu, Leonard Lipovich, Vladimir A Kuznetsov, Paul Robson, Lawrence W Stanton, Chia-Lin Wei, Yijun Ruan, Bing Lim, and Huck-Hui Ng. The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*, 38(4):431–440, Apr 2006.

[77] Qi Long Ying, Jennifer Nichols, Ian Chambers, and Austin Smith. Bmp induction of id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with stat3. *Cell*, 115(3):281–292, Oct 2003.

[78] Atsushi Suzuki, Angel Raya, Yasuhiko Kawakami, Masanobu Morita, Takaaki Matsui, Kinichi Nakashima, Fred H Gage, Concepcion Rodriguez-Esteban, and Juan Carlos Izpisua Belmonte. Maintenance of embryonic stem cell pluripotency by nanog-mediated reversal of mesoderm specification. *Nat Clin Pract Cardiovasc Med*, 3 Suppl 1:S114–S122, Mar 2006.

[79] Hitoshi Niwa. How is pluripotency determined and maintained? *Development*, 134(4):635–46, February 2007.

[80] Yuin-Han Loh, Weiwei Zhang, Xi Chen, and Joshy George. Jmjd1a and jmjd2c histone h3 lys 9 demethylases regulate self-renewal in embryonic stem cells. October 2007.

[81] Akitsu Hotta and James Ellis. Retroviral vector silencing during iPS cell induction: An epigenetic beacon that signals distinct pluripotent states. *Journal of Cellular Biochemistry*, 105(4):940–948, 2008.

[82] Bridget Keenen and Ivana L de la Serna. Chromatin remodeling in embryonic stem cells: regulating the balance between pluripotency and differentiation. *J Cell Physiol*, 219(1):1–7, Apr 2009.

[83] Katharine L Arney and Amanda G Fisher. Epigenetic aspects of differentiation. *J Cell Sci*, 117(Pt 19):4355–4363, Sep 2004.

[84] C. Francastel, D. Schuebeler, D. I. Martin, and M. Groudine. Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol*, 1(2):137–143, Nov 2000.

[85] Mark G Carter, Alexei A Sharov, Vincent VanBuren, Dawood B Dudekula, Condie E Carmack, Charlie Nelson, and Minoru S H Ko. Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol*, 6(7):R61, 2005.

[86] Dov Zipori. The nature of stem cells: state rather than entity. *Nat Rev Genet*, 5(11):873–878, Nov 2004.

[87] Melany Jackson, Anna Krassowska, Nick Gilbert, Timothy Chevassut, Lesley Forrester, John Ansell, and Bernard Ramsahoye. Severe global dna hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol*, 24(20):8862–8871, Oct 2004.

[88] Diana L Carlone, Jeong-Heon Lee, Suzanne R L Young, Erika Dobrota, Jill Sergesketter Butler, Joseph Ruiz, and David G Skalnik. Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking cpg binding protein. *Mol Cell Biol*, 25(12):4881–4891, Jun 2005.

[89] A. G. Smith, J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336(6200):688–690, Dec 1988.

[90] R. L. Williams, D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200):684–687, Dec 1988.

[91] H. Niwa, T. Burdon, I. Chambers, and A. Smith. Self-renewal of pluripotent embryonic stem cells is mediated via activation of stat3. *Genes Dev*, 12(13):2048–2060, Jul 1998.

[92] T. Matsuda, T. Nakamura, K. Nakao, T. Arai, M. Katsuki, T. Heike, and T. Yokota. Stat3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J*, 18(15):4261–4269, Aug 1999.

[93] Peter Cartwright, Cameron McLean, Allan Sheppard, Duane Rivett, Karen Jones, and Stephen Dalton. Lif/stat3 controls es cell self-renewal and pluripotency by a myc-dependent mechanism. *Development*, 132(5):885–896, Mar 2005.

[94] J. Wang, L. Y. Xie, S. Allan, D. Beach, and G. J. Hannon. Myc activates telomerase. *Genes Dev*, 12(12):1769–1774, Jun 1998.

[95] Rohan K Humphrey, Gillian M Beattie, Ana D Lopez, Nathan Bucay, Charles C King, Meri T Firpo, Stefan Rose-John, and Alberto Hayek. Maintenance of pluripotency in human embryonic stem cells is stat3 independent. *Stem Cells*, 22(4):522–530, 2004.

[96] T. Burdon, C. Stracey, I. Chambers, J. Nichols, and A. Smith. Suppression of shp-2 and erk signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol*, 210(1):30–43, Jun 1999.

[97] Ho Bum Kang, Jin Sook Kim, Hyung-Joo Kwon, Ki Hoan Nam, Hyun Soo Youn, Dai-Eun Sok, and Younghee Lee. Basic fibroblast growth factor activates erk and induces c-fos in human embryonic stem cell line mizhes1. *Stem Cells Dev*, 14(4):395–401, Aug 2005.

[98] Petr Dvorak, Dana Dvorakova, Stanislava Koskova, Martina Vodinska, Miroslava Najvirtova, Daniel Krekac, and Ales Hampl. Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells*, 23(8):1200–1211, Sep 2005.

[99] Mark E Levenstein, Tenneille E Ludwig, Ren-He Xu, Rachel A Llanas, Kaitlyn VanDenHeuvel-Kramer, Daisy Manning, and James A Thomson. Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem Cells*, 24(3):568–574, Mar 2006.

[100] C. Xu, M. S. Inokuma, J. Denham, K. Golds, P. Kundu, J. D. Gold, and M. K. Carpenter. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol*, 19(10):971–974, Oct 2001.

[101] M. Amit, M. K. Carpenter, M. S. Inokuma, C. P. Chiu, C. P. Harris, M. A. Waknitz, J. Itskovitz-Eldor, and J. A. Thomson. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol*, 227(2):271–278, Nov 2000.

[102] Petr Dvorak and Ales Hampl. Basic fibroblast growth factor and its receptors in human embryonic stem cells. *Folia Histochem Cytobiol*, 43(4):203–208, 2005.

[103] Ren-He Xu, Ruthann M Peck, Dong S Li, Xuezhu Feng, Tenneille Ludwig, and James A Thomson. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Meth*, 2(3):185–190, March 2005.

[104] Guangwen Wang, Hong Zhang, Yang Zhao, Jian Li, Jun Cai, Peigang Wang, Sha Meng, Jingbo Feng, Chenglin Miao, Mingxiao Ding, Dongsheng Li, and Hongkui Deng. Noggin and bfgf cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers. *Biochem Biophys Res Commun*, 330(3):934–942, May 2005.

[105] Chunhui Xu, Elen Rosler, Jianjie Jiang, Jane S Lebkowski, Joseph D Gold, Chris O'Sullivan, Karen Delavan-Boorsma, Michael Mok, Adrienne Bronstein, and Melissa K Carpenter. Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells*, 23(3):315–323, Mar 2005.

[106] Boris Greber, Hans Lehrach, and James Adjaye. Control of early fate decisions in human es cells by distinct states of tgfbeta pathway activity. *Stem Cells and Development*, 17(6):1065 – 1077, 2008.

[107] Xiaoxia Qi, Teng-Guo Li, Jing Hao, Jie Hu, Jing Wang, Holly Simmons, Shigeto Miura, Yuji Mishina, and Guang-Quan Zhao. Bmp4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A*, 101(16):6027–6032, Apr 2004.

[108] Ren-He Xu, Xin Chen, Dong S Li, Rui Li, Gregory C Addicks, Clay Glennon, Thomas P Zwaka, and James A Thomson. Bmp4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol*, 20(12):1261–1264, Dec 2002.

[109] Gillian M Beattie, Ana D Lopez, Nathan Bucay, Andrew Hinton, Meri T Firpo, Charles C King, and Alberto Hayek. Activin a maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cells*, 23(4):489–495, Apr 2005.

[110] Daylon James, Ariel J Levine, Daniel Besser, and Ali Hemmati-Brivanlou. Tgfbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*, 132(6):1273–1282, Mar 2005.

[111] Boris Greber, Hans Lehrach, and James Adjaye. Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human escs (hescs) to support hesc self-renewal. *Stem Cells*, 25(2):455–464, February 2007.

[112] J. A. Chong, J. Tapia-Ramirez, S. Kim, J. J. Toledo-Aral, Y. Zheng, M. C. Boutros, Y. M. Altshuller, M. A. Frohman, S. D. Kraner, and G. Mandel. Rest: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*, 80(6):949–957, Mar 1995.

[113] Nurit Ballas, Christopher Grunseich, Diane D Lu, Joan C Speh, and Gail Mandel. Rest and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell*, 121(4):645–657, May 2005.

[114] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*, 22(3):437–467, Mar 1969.

[115] J. MONOD, J. WYMAN, and J. P. CHANGEUX. On the nature of allosteric transitions: A plausible model. *J Mol Biol*, 12:88–118, May 1965.

[116] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys J*, 11(1):11–27, Jan 1971.

[117] S. Kauffman. Homeostasis and differentiation in random genetic control networks. *Nature*, 224(5215):177–178, Oct 1969.

[118] Stuart Kauffman, Carsten Peterson, Bjoern Samuelsson, and Carl Troein. Random boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25):14796–14799, December 2003.

[119] S. Huang. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *J Mol Med*, 77(6):469–480, Jun 1999.

[120] Hidde de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1):67–103, 2002.

[121] Z. Szallasi and S. Liang. Modeling the normal and neoplastic cell cycle with "realistic boolean genetic networks": their application for understanding carcinogenesis and assessing therapeutic strategies. *Pac Symp Biocomput*, pages 66–76, 1998.

[122] S. A. Kauffman. Antichaos and adaptation. *Sci Am*, 265(2):78–84, Aug 1991.

[123] Luis Mendoza and Ioannis Xenarios. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor Biol Med Model*, 3:13, 2006.

[124] Alessandro Di Cara, Abhishek Garg, Giovanni De Micheli, Ioannis Xenarios, and Luis Mendoza. Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics*, 8(1):462, 2007.

[125] A. E. Gelfand and C. C. Walker. On the character of an distance between states in a binary switching net. *Biol Cybern*, 43(2):79–86, 1982.

[126] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10):770–780, Oct 2008.

[127] Vijay Chickarmane, Carl Troein, Ulrike A Nuber, Herbert M Sauro, and Carsten Peterson. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol*, 2(9):e123, 2006.

[128] Ryo Matoba, Hitoshi Niwa, Shinji Masui, Satoshi Ohtsuka, Mark G Carter, Alexei A Sharov, and Minoru S H Ko. Dissecting oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS One*, 1:e26, 2006.

[129] Istvan Albert, Juilee Thakar, Song Li, Ranran Zhang, and Reka Albert. Boolean network simulations for life scientists. *Source Code Biol Med*, 3:16, 2008.

[130] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Juergen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi–a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, Dec 2006.

[131] Linda Petzold. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148, 1983.

[132] Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, March 2000.

[133] Shinji Masui, Yuhki Nakatake, Yayoi Toyooka, Daisuke Shimosato, Rika Yagi, Kazue Takahashi, Hitoshi Okochi, Akihiko Okuda, Ryo Matoba, Alexei A Sharov, Minoru S H Ko, and Hitoshi Niwa. Pluripotency governed by sox2 via regulation of oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*, 9(6):625–635, Jun 2007.

[134] James Ellis. Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene Ther*, 16(11):1241–1246, Nov 2005.

[135] Robert L Judson, Joshua E Babiarz, Monica Venere, and Robert Blelloch. Embryonic stem cell-specific micrornas promote induced pluripotency. *Nat Biotechnol*, 27(5):459–461, May 2009.

[136] Alexander Marson, Stuart S Levine, Megan F Cole, Garrett M Frampton, Tobias Brambrink, Sarah Johnstone, Matthew G Guenther, Wendy K Johnston, Marius Wernig, Jamie Newman, J. Mauro Calabrese, Lucas M Dennis, Thomas L Volkert, Sumeet Gupta, Jennifer Love, Nancy Hannett, Phillip A Sharp, David P Bartel, Rudolf Jaenisch, and Richard A Young. Connecting microrna genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, 134(3):521–533, Aug 2008.

[137] Masashi Kobayashi, Tatsuyuki Takada, Kentaro Takahashi, Yoichi Noda, and Ryuzo Torii. Bmp4 induces primitive endoderm but not trophectoderm in monkey embryonic stem cells. *Cloning Stem Cells*, 10(4):495–502, Dec 2008.

[138] Gudrun Valdimarsdottir and Christine Mummery. Functions of the tgfbeta superfamily in human embryonic stem cells. *APMIS*, 113(11-12):773–789, 2005.

[139] Zheng Li and Christina Chan. Extracting novel information from gene expression data. *Trends Biotechnol*, 22(8):381–383, Aug 2004.