Concise Review: The Dynamics of Induced Pluripotency and Its Behavior Captured in Gene Network Motifs

MAURO J. MURARO, HERMANNUS KEMPE, PERNETTE J. VERSCHURE

Swammerdam Institute for Life Sciences, Netherlands Institute for Systems Biology, University of Amsterdam, Amsterdam, The Netherlands.

Key Words. Induced pluripotent stem cells • Reprogramming dynamics • Systems behavior • Network motifs

ABSTRACT
The flexibility of cellular identity is clearly demonstrated by the possibility to reprogram fully differentiated somatic cells to induce pluripotent stem (iPS) cells through forced expression of a set of transcription factors. The generation of iPS cells is of great interest as they provide a tremendous potential for regenerative medicine and an attractive platform to investigate pluripotency. Despite having gathered much attention, the molecular details and responsible gene regulatory networks during the reprogramming process are largely unresolved. In this review, we analyze the sequence and dynamics of reprogramming to construct a timeline of the molecular events taking place during induced pluripotency. We use this timeline as a road map to explore the distinct phases of the reprogramming process and to suggest gene network motifs that are able to describe its systems behavior. We conclude that the gene networks involved in reprogramming comprise several feedforward loops combined with autoregulatory behavior and one or more AND gate motifs that can explain the observed dynamics of induced pluripotency. Our proposed timeline and derived gene network motif behavior could serve as a tool to understand the systems behavior of reprogramming and identify key transitions and/or transcription factors that influence somatic cell reprogramming. Such a systems biology strategy could provide ways to define and explore the use of additional regulatory factors acting at defined gene network motifs to potentially overcome the current challenges of inefficient, slow, and partial somatic cell reprogramming and hence set ground of using iPS cells for clinical and therapeutic use.


INTRODUCTION
Induced pluripotent stem cells (iPS cells) have gathered much attention since their discovery. iPS cells are a possible source of patient-specific stem cells making them very useful for regenerative medicine [1, 2]. iPS cells can also be used as a platform to model diseases, perform preliminary treatment screens, and analyze pluripotency, development, and differentiation [3, 4]. Improvements in reprogramming and selection techniques provided iPS cells that are increasingly similar to embryonic stem cells (ESCs) regarding their morphological characteristics as well as their genome-wide transcriptional activity and epigenetic composition [5–7] (for review [8]). Reprogramming has been achieved with various different animal species and starting from different somatic cell types. In principle, a variety of reprogramming methods exist, using different types of (integrating and nonintegrating) viruses, proteins [9], episomal vectors [10], and even (micro)RNAs [11] (for review [12]).

There are several challenges related to reprogramming somatic cells into iPS cells. First, the reprogramming efficiency is usually very low (~0.1% with retrovirus to ~10% with RNA), which makes it difficult to isolate or follow the reprogrammed cells for analysis and to obtain homogenous cell populations. Second, during reprogramming, colonies appear that resemble ESCs, which are not completely reprogrammed. Third, direct reprogramming is slow (taking days or weeks), especially when compared with other methods of nuclear reprogramming such as cell fusion experiments [13, 14]. Two different models, the deterministic and stochastic model, have been proposed to explain the low efficiency and slow kinetics of reprogramming [15]. A recent study used a stochastic model to simulate a probabilistic Boolean network describing the interplay between changes in gene expression and induced changes in the epigenetic composition [16]. Clearly, extensive understanding of the molecular mechanisms behind the reprogramming process is essential in order to establish safe and efficient ways to generate therapeutic iPS cells that are applicable for therapeutic use [17]. Here, we construct a chronological timeline of the direct reprogramming process and we determine the nature of the different phases of reprogramming (Table 1; Fig. 1). Using information from this constructed timeline, we explore possible explanations for the various observations during the reprogramming process and we suggest gene network motifs that are able to describe systems behavior of this dynamic process. We conclude that the gene networks involved in reprogramming comprise several feedforward loops combined with autoregulatory behavior and other regulatory network motifs that can...
explain the observed dynamics of induced pluripotency (for review [25]). The behavior of such gene network motifs is described in paragraph “Gene network motifs” and Figure 2 and applied to reprogramming in Figure 3. Our constructed timeline and proposed gene network motif behavior could serve as a guideline to study and advance somatic cell reprogramming. In this context, we highlight the usefulness of exposure to additional regulatory factors exhibiting for instance signaling, tumor suppressor, and/or epigenetic modifying characteristics (Table 2) to improve the reprogramming process and thus provide ways for therapeutic use of iPS cells.

The most popular choice in iPS studies, since mouse cells are generally more available, have faster reprogramming dynamics, and have higher reprogramming efficiency. Several types of measurements are used to characterize the reprogramming cells, including changes in cellular morphology, fluorescence-activated cell sorting analysis of the presence of cell surface markers, analysis of pluripotency gene expression levels, and the epigenetic composition [8]. Depending on the studies performed, different reprogramming efficiencies and dynamics are observed. The time points indicated in the figures and text are for reference, and should not be regarded as fixed.

Early Changes
Direct reprogramming starts with the expression of exogenous transgenes, RNA, or protein, after which the starting cell population begins to reprogram into iPS cells. Rapid, genome-wide changes in the histone modification composition such as dimethylation of histone H3 at lysine 4 (H3K4me2) and

TIMELINE OF INDUCED PLURIPOTENCY

Several groups have studied the molecular events that take place during reprogramming (Table 1). Mouse cells have been the most popular choice in iPS studies, since mouse cells are generally more available, have faster reprogramming dynamics, and have higher reprogramming efficiency. Several types of measurements are used to characterize the reprogramming cells, including changes in cellular morphology, fluorescence-activated cell sorting analysis of the presence of cell surface markers, analysis of pluripotency gene expression levels, and the epigenetic composition [8]. Depending on the studies performed, different reprogramming efficiencies and dynamics are observed. The time points indicated in the figures and text are for reference, and should not be regarded as fixed.

Early Changes
Direct reprogramming starts with the expression of exogenous transgenes, RNA, or protein, after which the starting cell population begins to reprogram into iPS cells. Rapid, genome-wide changes in the histone modification composition such as dimethylation of histone H3 at lysine 4 (H3K4me2) and
Figure 1. Dynamics of the molecular events during reprogramming. (A): Somatic cell reprogramming consists of two general phases. The first phase exhibits transgene dependence: if exogenous gene expression is removed, cells revert to a parent somatic cell type. The latter part of the dependent phase is typical for pre-induced pluripotent stem (iPS) or partially reprogrammed cells, which can be maintained as a stable state in culture. The second phase is initiated by independence from transgenes and by activation of the endogenous pluripotency circuitry. While the dependent phase has a determined sequence of molecular events, the independent phase is more gradual in nature, consisting of various epigenetic resetting events that may continue after the establishment of early passaged iPSCs. (B): A set of morphological changes are noticed during the dependent phase. Successfully reprogrammed cells typically have a reduced cell size and proliferate faster, in principle from their first cell division. Faster cell division has been attributed to a shorter G1 phase in the cell cycle of reprogramming cells. These cells create clusters of reprogramming cells that then form microcolonies and finally form colonies that activate endogenous pluripotency genes and exhibit ESC-like morphology. These colonies can be subcloned and expanded in culture. (C): During the reprogramming process defined sets of genes switch their transcriptional activity. Downregulation of somatic genes like Thy1 and upregulation of proliferative genes and a mesenchymal-to-epithelial transition illustrate the first set of transcriptional changes. This is followed by a sequential upregulation of embryonic genes such as AP, SSEA1, and Fbx15. Lastly, endogenous pluripotency-related genes like Oct4, Sox2, and Nanog are activated. Exogenous transgenes start to be silenced early in the pre-iPS cell phase of reprogramming and continue to be silenced into the independent phase. (D): A set of chromatin remodeling events are noted during the reprogramming process: the earliest described changes in the chromatin composition are the resetting of histone methylation patterns in enhancer and pluripotency gene promoters. The histone mark H3K27me3 that is positioned on bivalent chromatin regions is removed from pluripotency promoters and replaced by H3K4me2. During the transgene-independent phase of reprogramming, several gradual chromatin remodeling processes take place, such as X-chromosome reactivation (in mouse), telomere elongation, epigenetic pattern shifts, and decrease of DMRs. Abbreviations: DMT, differentially methylated regions; ESC, embryonic stem cell.
Figure 2. Behavior of several gene network motifs. Here, we show an overview of the gene network motifs that can explain parts of the observed behavior of induced pluripotency. The gene network motifs and their behavior over time are shown in the left column of the image, the colors used in the motifs correspond to the colors used in the time plots. In the right panels, the behavior of these motifs is summarized. Panel (A) shows positive auto regulation and panel (B) shows double negative regulation both motifs can lead to bistable behavior. Panel (C) illustrates a cascade and panel (D) shows the AND gate motif and are both able to cause a delay between the signal and response. In some examples, we varied either the strength or the length of the signal to show characteristics of the motif. In most cases, similar behavior can be achieved by varying the signal strength or length, since the behavior of the motif is always dependent on a reactant to pass a threshold value. The signal (S) is active for the indicated time (gray areas) the signal strength is 1, except if indicated differently. The $h_a(X)$ and $h_r(X)$ describe the steady-state response of gene activation or repression given X and are modeled with a Hill function. The parameter $a$ is the maximal activation or repression that can be achieved, $k$ determines for which $X$ the response is half maximal, and $n$ sets the steepness of the response curve. For high $n$, this Hill function approximates a logic function, which can be either on or off. For a more extensive and thorough description of this matter we refer to [25–28].

Positive auto regulation
Species X stimulates its own expression. When this feedback is cooperative two distinct steady states can be reached by the system under the same conditions. Switching between both steady states can be achieved by an external interference in the system. In this box the system is shown for two different signal strengths ($S=1$ and $S=3$), showing that a strong signal can induce an enhanced steady state (dark blue line)

$$\frac{dx}{dt} = S \cdot k_x + h_a(X) \cdot k_{px} - X \cdot k_{dx}$$

Parameters: $k_x=k_{px}=k_{dx}=1, \alpha=3, k=1.5, n=5$

Properties: Bistable

Double negative regulation
Factors X and Y have an inhibitory effect on each other’s production. This wiring can effect mutually exclusive expression. Switching between the two steady states is achieved by external interference, which disturbs the balance in the system. In this box the production of X is temporally stimulated by the signal to induce the steady state switch.

$$\frac{dx}{dt} = h_r(Y) \cdot k_{px} - X \cdot k_{dx} + S \cdot k_x, \quad \frac{dy}{dt} = h_a(X) \cdot k_{py} - Y \cdot k_{dy}$$

Parameters: $k_x=k_{dx}=k_{dy}=k_{dy}=1, \alpha=2, k=1, n=5$

Properties: Bistable, mutually exclusive

Cascades
In a cascade design factor X stimulates the production of downstream factor Y while Y stimulates the production of factor Z. This wiring causes a delay between the occurrence of the stimulus and the response in Z.

$$\frac{dx}{dt} = S \cdot k_x - X \cdot k_{dx}, \quad \frac{dy}{dt} = h_a(X) \cdot k_{py} - Y \cdot k_{dy}$$

$$\frac{dz}{dt} = h_a(Y) \cdot k_{pz} - Z \cdot k_{dz}$$

Parameters: $k_x=4, k_{dz}=k_{dy}=k_{pz}=k_{ps}=1, k_{dx}=k_{dy}=1, \alpha=2, k=1, n=5$

Properties: Delay

AND gate
In contrast to the cascade design, in the AND gate the activation of Z does not occur when the stimulus is terminated. This setup gives a delay in Z production and additionally acts as a strong filter for transient signals.

$$\frac{dy}{dt} = h_a(S) \cdot k_{py} - Y \cdot k_{dy}$$

$$\frac{dz}{dt} = h_a(S) \cdot h_a(Y) \cdot k_{pz} - Z \cdot k_{dz}$$

Parameters: $k_{dy}=1, k_{dy}=k_{dz}=k_{dy}=0.5, k_{pz}=k_{ps}=1, k_{dx}=k_{dy}=1, \alpha=2, n=5, k=1.5, n=100$

Properties: Delay, Transient signal filter
Figure 3. Gene network structure explaining delayed reprogramming dynamics and low efficiency. Reprogramming can be described by a set of gene network motifs. The properties and behavior of such motifs is as outlined in more detail in Figure 2A–2D. (A): A combination of the feedforward loop, cascade, and AND gate gene regulatory network motifs can explain the behavior of induced pluripotency. The activation of endogenous NOS expression is dependent on input from several feedforward loops and of input of exogenous OSKM expression, relayed to an AND gate motif. This motif requires both inputs to generate output (activation of endogenous NOS expression). If one of the two inputs is missing or removed before the initiation signal is relayed through the feedforward loops, there is no output signal. (B): Module behavior representation of the feedforward loop, cascade, and AND gate motif. If the initiation signal (exogenous OSKM expression) is of short duration, it will activate genes directly regulated by OSKM as well as the earliest pluripotency genes like AP. As these changes take some time, a second group of pluripotency genes (SSEA1) is activated later and it takes some time for the initiation signal to be relayed to the AND gate. If the signal disappears before this point, the AND gate will receive input from the SSEA1 group of genes, but not from the exogenous OSKM. Therefore, the system will not change (i.e., switch to the production of endogenous NOS thereby setting-off for successful reprogramming) if the signal duration is short (in principle transgene dependence is noted until the intermediate pre-iPS state, ~day 7). In case the signal duration is too short to provide induction of endogenous NOS expression and hence efficient reprogramming, a couple of factors have been noted to be able to allow switching into the efficient reprogramming. Faster gene expression in response to chromatin remodeling can lead to a faster response through the cascade, leading to more efficient reprogramming. An example of such change is shown with the red and orange dashed line, where the critical concentration of SEAA1 (dotted black line) to switch-on endogenous NOS is reached during the signal event. Abbreviation: OSKM, Oct4, Sox2, Klf4, and c-Myc.

Table 2. Overview of factors known to affect reprogramming dynamics

<table>
<thead>
<tr>
<th>Type</th>
<th>Treatment/factor</th>
<th>Supposed role in reprogramming</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor suppressor</td>
<td>P53-P21 silencing</td>
<td>Reprogramming accelerated (early phase) through increased proliferation</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>Ink4/Arf locus silencing</td>
<td>Reprogramming accelerated (early phase) through increased proliferation</td>
<td>[30]</td>
</tr>
<tr>
<td>Chromatin modifier</td>
<td>WDR5</td>
<td>H3K4 methylation mediated (early phase)</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>DNA methyltransferase inhibitor (5-Aza-C)</td>
<td>Reprogramming accelerated (late phase); pluripotent epigenome established</td>
<td>[23]</td>
</tr>
<tr>
<td></td>
<td>DNMT3a 3b knockout</td>
<td>iPSC without developmental potential</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>BAF complex components</td>
<td>Chromatin remodeling</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td>AID</td>
<td>Key pluripotency gene promoters demethylated</td>
<td>[14]</td>
</tr>
<tr>
<td></td>
<td>Histone deacetylase inhibitor</td>
<td>Reprogramming accelerated</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td>PRC2 subunit overexpression</td>
<td>Reprogramming efficiency accelerated through silencing of lineage-associated genes</td>
<td>[35]</td>
</tr>
<tr>
<td>Signaling molecule</td>
<td>BMP</td>
<td>MET transition accomplished</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td>Wnt</td>
<td>Reprogramming accelerated</td>
<td>[37]</td>
</tr>
<tr>
<td></td>
<td>Stat3 activation</td>
<td>Progression to ground state pluripotency limited</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>Nanog overexpression</td>
<td>Reprogramming accelerated through proliferation-independent fashion</td>
<td>[39]</td>
</tr>
</tbody>
</table>

Overview of a set of regulatory factors to have been identified to improve the reprogramming dynamics and/or efficiency acting as tumor suppressors, as regulators of the epigenetic composition, or acting through signaling cascades. The supposed role of the regulatory factors and their reference are given. Abbreviations: AID, activation-induced cytidine deaminase; iPSC, induced pluripotent stem cell; MET, mesenchymal-to-epithelial.
trimethylation of histone H3 at lysine 27 (H3K27me3) have been reported at more than a thousand loci, including pluripotency-related or developmentally regulated gene promoters, that is, Sox2, Klf4, and c-Myc that are often significantly enriched in Oct4 and Sox2 protein binding sites [40]. Moreover, a change in H3K4me on enhancer regions is noted (Fig. 1), thereby marking the changes of a somatic cell signature to an ESC-like signature during this time frame which continues during later cell culture phases [40, 41]. Transcription from the pluripotency-related loci does not take place until later in the reprogramming process, as proven by the enrichment of H3K36me3 and RNA PolII, suggesting that these genes are at this point only primed for transcription [40]. The iPS reprogramming is characterized by exogenous activation of the pluripotency network which subsequently autoregulates its endogenous expression [42, 43].

Morphology and Proliferation
Small, rounded cells and increased proliferation (detected within 2–3 days of reprogramming factor induction) are obvious characteristics of ESCs (Table 1). Live imaging experiments showed that cells that undergo faithful reprogramming show increased proliferation and reduced cell size even before the first cell division [21]. Not surprisingly, genes involved in DNA replication (Poli, Rfc4, and Mcm5) and cell cycle progression (Ccn1 and Ccn2) are upregulated during the early phase of direct reprogramming, which has been linked to the proliferation-promoting role of c-Myc [23]. After the change in cell morphology, cells start forming microcolonies (around day 3–4), which continue to grow into large, ES-like iPS cell colonies around day 7–10 (Table 1; Fig. 1).

Mesenchymal-to-Epithelial Transition and Transcriptional Changes
Early transcriptional changes concern the downregulation of lineage-specific genes such as fibroblast markers Thy-1 and Sna1 in mouse and CD13 in human fibroblasts (Table 1). Many of these downregulated genes are mesenchymal in nature, and their downregulation is accompanied by upregulation of epithelial genes, marking a mesenchymal-to-epithelial transition (MET) within reprogramming fibroblasts [24, 44], which is the opposite reprogramming direction of an epithelial-to-mesenchymal transition observed during differentiation [24, 44]. Other genes that increase in expression during this phase are stress-induced and antiproliferative genes, often part of tumor suppression pathways [23]. These observations indicate that transcription factor induction might trigger mechanisms that normally prevent uncontrolled proliferation and thus block or delay cells during the reprogramming process. Typically following the downregulation of lineage markers, early pluripotency genes are upregulated (Fig. 1). Of these early pluripotency genes, alkaline phosphatase is activated first, followed by activation of the embryonic surface factor SSEA1 and the ES marker gene Fbx15, all preceding activation of the endogenous key pluripotency genes like Nanog and Oct4 both acting during the later phase of reprogramming [22] (Table 1).

Transgene silencing, a hallmark of pluripotency, starts as early as day 4 and gradually continues after activation of the endogenous pluripotency circuitry until transgene levels are completely silenced. Typically, at day 14 of reprogramming, only less than 5% of the cells that have activated their endogenous stemness genes and embryonic surface factor SSEA1 still show detectable transgene levels [18]. It is noteworthy that iPS cells can be obtained with lentiviruses [45] that are continuously active. However, the differentiation potential of iPS cells seems to benefit from the excision of transgenes after completing the reprogramming process [45, 46].

Transgene Independence
The transition from the early phase to the late phase of reprogramming is marked by the independence toward exogenous reprogramming factor expression (Fig. 1). Exogenous expression of the pluripotency-induced transcription factors is required for 7–12 days in mouse cells [18, 20, 22, 47]. Human cells need longer than 10 days of exogenous reprogramming factor expression [48]. If transgene expression is induced for a period shorter than a week, existing colonies usually regress within 2 days to a somatic cell phenotype, indicating a clear flexibility in the reprogramming process [18, 22, 24]. It should be noted that although transgene-independent iPS cells arise after 7 days, continuing transgene expression for up to 3 weeks further enhances reprogramming efficiency [22, 24]. This can be explained by the stochasticity of the direct reprogramming process: some cells reach transgene independence at later time points than the first reprogrammed cells [39].

To summarize, the early phase of reprogramming starts with an increased proliferation rate, decrease in cell size, and fast changes in the histone methylation pattern (H3K4me2 and H3K27me3) at promoter and enhancer regions of genes expressing transcription factors that are crucial for the reprogramming process. Subsequent steps of the early phase include a transcriptional shift from differentiated state, downregulation of lineage-specific markers, a MET, and the upregulation of early pluripotency-related genes such as alkaline phosphatase and the embryonic surface marker SSEA1 (Fig. 1).

Late Phase
The early phase ends when reprogrammed cells become independent of their exogenous transcription factor expression typically activating the endogenous pluripotency circuitry (Table 1; Fig. 1) [43]. Endogenous pluripotency genes are detected after a week of direct reprogramming in mouse fibroblasts by the activation of endogenous Oct4, Sox2, and other pluripotency-related genes such as Nanog, Rex1, and Essrb [18]. The activation of transcription factor Nanog has been directly linked to the reprogramming commitment [24] and it is essential for reaching a true pluripotent ground state [49]. Selection for Nanog as well as for Oct4 is shown to deliver more iPS cells that are faithfully reprogrammed compared to selecting for an earlier reprogramming marker gene such as Fbx15 [5–7, 50]. It should be noted, however, that Oct4 activation can also occur in partially reprogrammed colonies [51]. The beginning of the late phase in reprogramming coincides with the appearance of the first large ES-like colonies.

X Chromosome Reactivation
X chromosome reactivation during reprogramming is a late event that is initiated after the activation of endogenous genes [18, 52]. Direct reprogramming of mouse cells is accompanied by a faithful reactivation of the silenced X chromosome and its subsequent random inactivation [5]. Human cells, however, do not reactivate the inactive X chromosome under the same conditions. Some human iPS cells do activate the inactive X chromosome at low frequency, something that seems to be dependent on derivation conditions such as the use of feeder cells [53]. This observed difference in human iPS versus mouse ESCs is reprogramming is thought to be caused by the difference in the pluripotency state of human versus mouse, the human iPS and ESCs being in a developmentally more advanced pluripotent state (the primed state),
compared to mouse stem cells being in a rather naïve pluripotent state when not treated with more specific culture conditions [5, 6, 50, 54, 55].

**Telomere Elongation**

Another important characteristic of ESCs is their ability for indefinite self-renewal. Indispensable for this process is telomerase reverse transcriptase (tet)-mediated telomere protection. Activation of tet during the reprogramming process has been shown early during the reprogramming process [5, 56, 57]. The actual telomere elongation during reprogramming is a more time-consuming process than the activation of tet. Early passage iPSCs have telomeres exhibiting an intermediate length when compared with those of ESCs, but this difference disappears when cell passaging continues [57]. Importantly, telomere elongation is also observed in iPSCs cells derived from aged animals, suggesting that any faithfully reprogrammed cell gains the telomere length and immortality typically associated with ESCs. In principle, telomeres progressively shorten again during the subsequent differentiation of iPSCs to fibroblasts [58]. Intriguingly, in iPSCs cells, telomeric heterochromatin contains the same histone methylation pattern (H3K9me3 and H3K4me3) and subtelomeric DNA hypermethylation pattern as observed in ESCs. Such patterns are a typical characteristic of ESCs [57, 58] (Fig. 1).

**Global Changes in the Epigenetic Composition**

The first changes in the epigenetic composition take place in the early reprogramming phase on the promoters of pluripotency genes (Fig. 1) that are activated after the first week of reprogramming. During the late phase of reprogramming, a more gradual change in DNA and histone methylation takes place.

Bivalent domains are found on promoters of developmentally important genes and their bivalent epigenetic composition is thought to enable silencing of these developmental genes in ESCs while keeping them poised for activation upon lineage specification [59–62]. Bivalent chromatin domains typically contain opposing epigenetic marks, that is, both transcriptionally active and repressive histone marks, H3K4me2 and H3K27me3, respectively. The genome-wide H3K4me3 and H3K27me status of promoters and bivalent chromatin domains in iPS and ESCs are nearly indistinguishable [23]. The DNA methylation state is another epigenetic mark of importance for the reprogramming outcome. In principle, CpG methylation represses gene activity by preventing binding of the transcription machinery and by recruiting enzymes that induce repressive histone modifications, for example, histone deacetylases [63]. In iPSCs, the promoters of genes belonging to the circuitry of pluripotency are DNA hypomethylated in an “ES cell-like manner” [5, 23, 56]. The dynamic changes in the DNA methylation pattern, that is, the differentially methylated regions (DMR) in iPSCs, are mainly found in CpG islands showing a memory of the somatic cell type that was used as reprogramming starting point [64]. These DMRs have been illustrated to be abundantly present in early passaged iPSC cells and to diminish during prolonged passaging [65]. Others have confirmed this erasure of the epigenetic memory upon prolonged cell passaging [66].

Epigenetic differences between ESCs and iPSCs are a subject of much debate, as they might indicate the existence of an epigenetic memory (epigenetic “traces” in iPSC cells reminiscent of the somatic cell type of origin). Although some studies showed that iPSC cells could be clustered according to their parent somatic cell type based on their gene expression profiles [67–69], others refuted this [70]. Further investigation on genome-wide H3K4/H3K27 methylation patterns and RNA and microRNA expression patterns of genetically matched ESCs and iPSCs cells proved that iPSC cells are indistinguishable from ESCs with one exception, that is, the Dlk1-Dio3 gene cluster that is generally imprinted during differentiation [20].

In short, the late phase of reprogramming encompasses several (epi)genetic-related changes: the activation of endogenous pluripotency-related regions, transgene silencing, X chromosome reactivation, telomere elongation, and the erasure of epigenetic memory. These late phase processes illustrate a less well sequentially ordered succession of events compared with the sequentially ordered changes taking place during the early phase. The late phase events have a more gradual character, needing continued cell passaging to establish cells that have all the ES-like characteristics. In conclusion, reprogramming occurs as a two-phase process. In the following section, we extrapolate our constructed time lining including the early and late phase observations discussing the systems behavior of reprogramming dynamics and proposing gene network motifs that can explain the observed behavior.

**SYSTEM PARAMETERS OF INDUCED PLURIPOTENCY**

The observations listed in Table 1 and Figure 1 show that there are two distinct phases in this process, that is, the exogenous-induced transcription factor-dependent and -independent phase. These observations raise a number of questions. For example: what triggers the irreversible switch from the transgene dependent to transgene-independent phase? Why is the sequence of genes activated during reprogramming sequentially determined during the early phase? And what explains the slow speed and inefficiency of reprogramming during induced pluripotency when compared with cell fusion nuclear reprogramming [14]? Capturing the dynamics of induced pluripotency in simple gene network motifs provides a way to define systems parameters that play a role in the induced somatic cell reprogramming process. Such a gene network motif representation offers tools to predict which parameters are most critical to fulfill successful reprogramming behavior. Here, we present a set of gene network motifs that allow to explain the behavior of the successive early and late phase reprogramming steps. Such gene network motifs could aid in identifying critical components to fulfill improved reprogramming thereby potentially setting ground to use iPSCs for clinical purposes.

**Gene Network Motifs**

The observed behavior in the reprogramming process can be matched with the behavior of regulation motif patterns regulation [5, 6, 50, 54, 55]. The dynamic changes in the DNA methylation pattern, that is, the differentially methylated regions (DMR) in iPSCs, are mainly found in CpG islands showing a memory of the somatic cell type that was used as reprogramming starting point [64]. These DMRs have been illustrated to be abundantly present in early passaged iPSC cells and to diminish during prolonged passaging [65]. Others have confirmed this erasure of the epigenetic memory upon prolonged cell passaging [66].

Epigenetic differences between ESCs and iPSCs are a subject of much debate, as they might indicate the existence of an epigenetic memory (epigenetic “traces” in iPSC cells reminiscent of the somatic cell type of origin). Although some studies showed that iPSC cells could be clustered according to their parent somatic cell type based on their gene expression profiles [67–69], others refuted this [70]. Further investigation on genome-wide H3K4/H3K27 methylation patterns and RNA and microRNA expression patterns of genetically matched ESCs and iPSCs cells proved that iPSC cells are indistinguishable from ESCs with one exception, that is, the Dlk1-Dio3 gene cluster that is generally imprinted during differentiation [20].

In short, the late phase of reprogramming encompasses several (epi)genetic-related changes: the activation of endogenous pluripotency-related regions, transgene silencing, X chromosome reactivation, telomere elongation, and the erasure of epigenetic memory. These late phase processes illustrate a less well sequentially ordered succession of events compared with the sequentially ordered changes taking place during the early phase. The late phase events have a more gradual character, needing continued cell passaging to establish cells that have all the ES-like characteristics. In conclusion, reprogramming occurs as a two-phase process. In the following section, we extrapolate our constructed time lining including the early and late phase observations discussing the systems behavior of reprogramming dynamics and proposing gene network motifs that can explain the observed behavior.
Delayed activation of downstream factors can be caused by a cascade design motif. Within a cascade, species $X$ stimulates production of downstream species $Y$, while $Y$ has to stimulate production of species $Z$. To stimulate production of downstream species, the regulator first has to reach a threshold value. The time to reach this threshold sets a delay between sequential activating species and slows down the response over the cascade (Fig. 2C). The same effect can be achieved with a gene network motif referred to as AND gate motif, in which the initial signal has to be present to activate the last species in the cascade (Fig. 2D). This setup not only gives a delayed response but also is a strong filter for transient signals. Below we explain how the observed interplay of reprogramming factors is matched with the regulation motif patterns described.

**Irreversible Switch Behavior**

Reprogramming cells are dependent on exogenous reprogramming factor expression for successful reprogramming up to the activation of the endogenous pluripotent gene circuitry. When the endogenous gene circuitry is established, cells irreversibly switch into the late, transgene-independent phase of reprogramming (Fig. 1). This reprogramming switch is most clearly shown in cell systems using doxycyclin inducible transgene expression. If doxycyclin is removed before ~day 7 of the reprogramming process, cells revert to a fibroblast state [22, 24]. After this time point, iPS colonies emerge, even in the absence of doxycyclin. In other words, the expression of the exogenous transgenes helps to “jump-start” the pluripotent gene circuitry [43]. After having reached the endogenous transgene-independent state, there is typically memory in the system and the cells remain in a stable, pluripotent state. Here, we outline several gene network motifs that are able to trigger such an irreversible switch and a memory state.

**Positive Autoregulation**

Oct4, Sox2, and Nanog (together referred to as NOS) with other pluripotency genes, represent the core pluripotent gene circuitry (Fig. 2A). Oct4, Sox2, Klf4, and c-Myc reprogramming factors (together referred to as OSKM) are able to exogenously trigger a switch in endogenous NOS levels thereby maintaining the reprogramming state. Similar to general transcription factors, the pluripotent gene circuitry transcription factors bind to thousands of promoters to repress transcription from somatic cell genes and initiate transcription from genes related to pluripotency and self-renewal, including their own expression by autoregulation [43, 71]. Positive autoregulation, if strong enough, can lead to a bimodal distribution of the produced protein levels. For the behavior of this positive autoregulation motif, we can for example consider a gene that regulates itself and is expressed only at a low level. When the expression of this gene is increased upon the act of a gene-related input signal, the corresponding protein levels of this gene will remain elevated, even after removal of the initial input signal [25]. The autoregulatory nature of Oct4, Sox2, and Nanog could represent such a motif inducing an ESC switch [72]. The gene-related input could initiate a switch in the levels of reprogramming factors Oct4 and Sox2. If the input signal, that is, exogenous OSKM expression (i.e., Oct4, Sox2, Klf4, and c-Myc), is a short pulse, the endogenous NOS levels (i.e., Oct4, Sox2, and Nanog), which are essential for maintaining pluripotency, will not reach a new steady-state level. However, when the signal is provided during a substantially extended period thereby reaching a switching threshold, endogenous NOS levels will remain stable, even after removal of the initial signal.

**Double Negative Autoregulation**

Another network motif to consider is that of double negative autoregulation, in which two groups of genes repress each other thereby evoking a switch-like gene network behavior [25] (Fig. 2B). As a simplified example during reprogramming, we can consider the mesenchymal and stemness genes. During the reprogramming process, mesenchymal genes are initially expressed, while endogenous pluripotency-related genes are not yet expressed. Exogenous expression of OSKM represses the fibroblast genes and activates endogenous NOS transcription, which then remains active through repression of somatic cell-specific genes and positive autoregulation [42]. In principle, this motif could lock the cell in a changed steady state, even after removal of exogenous OSKM. Again, the input signal plays an important role. The levels of endogenous NOS and of fibroblast genes will only reach and remain a new steady-state level if the expression of exogenous OSKM has a substantial timing. The fact that pluripotent genes repress somatic cell-type genes is clear from the observed rapid downregulation of mesenchymal marker genes observed in the early phase of the reprogramming process (Fig. 1). In support of this notion, pre-iPS cells often have incompletely repressed somatic cell genes [23]. In conclusion, the memory and switch-like behavior of a gene network can originate from the influence of autoregulatory positive feedback loops and likely from double negative feedback between groups of genes that are able to either activate or repress the pluripotent cell character.

**Slow Reprogramming Dynamics**

Exogenous OSKM expression is observed within 24 hours after infection with reprogramming factors and accordingly the first changes in the reprogramming process take place very fast. For example, the increase in proliferation occurs already after one cell division. Some pluripotency markers like alkaline phosphatase and SSEA1 are expressed later on (after 3–4 days), while the cells are still in the transgene-dependent phase. Endogenous NOS expression is not observed until ~day 7, at the phase that the formation of large ES-like colonies starts. Clearly these endogenous NOS levels are autoregulatory, but they are not immediately induced upon exogenous OSKM expression. We illustrate the network motifs that could infer this delayed expression of endogenous NOS.

**Signaling Cascade**

Since OSKM can bind a high number of targets either at the cluster and the individual level, the precise way in which the overexpression of these target genes triggers induced pluripotency is unclear (Fig. 2C). Transcription factors, somatic cell genes, and other pluripotency genes or regulators like Nanog are all possible OSKM targets [71]. Since the OSKM target connections are not all known, it is best to depict the genes affected by OSKM as a group of genes rather than as a set of individual genes. The genes in this group will in turn influence other groups of genes, forming a cascade of transcription activation that might take several cell generations to unfold [25]. Such a cascade gene network could explain the sequential character of the early reprogramming phase (Fig. 3). This, however, does not explain why the increased exogenous OSKM levels do not immediately induce endogenous NOS expression or that of other OSKM-regulated pluripotency genes. Nanog, another key pluripotency gene, is also activated late during the reprogramming process and Nanog is dispensable during the early phase of reprogramming [73]. These observations imply that the late pluripotency genes, despite their autoregulatory nature, depend on output from the
signaling cascade that can activate the early pluripotency genes. Recent studies measuring RNA transcripts in single cells indeed showed that the late phase of reprogramming is a hierarchical process, that is, Sox2 activating Sal4 and Lin28, followed by activation of other pluripotency regulator genes such as Fgf4 and Ptx1 [51].

AND Gate
Another gene motif that could explain the delay observed in endogenous NOS activation is the existence of an AND gate motif in the pluripotent circuitry (Fig. 3) (Fig. 2D). The AND gate is a network motif that receives input from two (or more) sources and has one output only if all inputs are present. In this case, the activation of endogenous NOS is dependent on input from several feedforward loops. If the initiation signal (exogenous OSKM levels) is of short duration, it will activate the genes directly regulated by OSKM (e.g., proliferative genes and alkaline phosphatase expression, which are some of the earliest transcriptional changes in reprogramming).

Table 1. As the activation of this group of genes takes some time, a second group of pluripotency genes (i.e., embryonic surface marker SSEA1) is activated later causing a delay before the initiation signal is relayed to the AND gate (Fig. 3). If the signal disappears before this point, the AND gate will have input from the SSEA1 group of genes, but not from the exogenous OSKM. Therefore, the system will not change if the signal duration (exogenous OSKM levels) is short (i.e., transgene dependence is maintained until day 7). Another candidate for input in the AND gate is the MET, which has been shown to be important during the early phase of reprogramming and shown to be regulated by bone morphogenetic protein (BMP) signaling. BMP signaling does not activate NOS on itself, but synergizes with OSKM to stimulate the onset of Nanog and Sal4 expression [24].

Low Efficiency
The network motifs described above can partly explain the observed reprogramming behavior, that is, positive feed forward loops in autoregulation, double negative feedback regulation, and AND gate motifs illustrating the slow dynamics of reprogramming and its switch-like behavior. However, the efficiency (or lack thereof) and variability of the reprogramming process exhibit performance that cannot be captured in the behavior of these gene network motifs. Such variable reprogramming can be explained by considering the opposed movement of development in the context of Konrad Waddington’s epigenetic landscape (for further reading [74]). The navigation through this noisy attractor landscape is likely to depend on stochastic fluctuations in the pluripotency inducing transcription factor levels, much like fluctuations in defined lineage choices during differentiation [74]. For example, the levels of SCF in blood progenitor cells have been shown to be correlated with normal differentiation in a cellular population. Choice of lineage was shown to be dependent on the levels of induction factors such as SCA-1 and Nanog [75, 76]. It is likely that the transitions through the sequential steps during the reprogramming process are determined by stochastic fluctuations in the levels of key transcription factors.

Taken together, the timeline observations from Table 1 and Figure 1 displaying the dynamics of induced pluripotency can be captured by gene network motif representations (Figs. 2, 3). We propose that systems network behavior and simulation of the dynamics of induced pluripotency by such gene network models be crucial to fully understand the mechanistic behavior of somatic cell reprogramming. Of course, the gene networks involved in in vivo reprogramming are much more complex than the networks described here. Still, such a simplified representation can give substantial mechanistic insight to pinpoint key switching steps of the reprogramming system, that is, a drop or increase in essential reprogramming factors, a switch in the permissiveness of the epigenetic state. Moreover, our constructed timeline that we use as input to outline gene network motif behavior of the reprogramming dynamics provides a basis to identify additional reagents based on their effect at key switching system steps to overcome the low and inefficient iPS cell production.

Factors that Accelerate Reprogramming Dynamics
A variety of factors that accelerate reprogramming dynamics have been identified in gene knockout and overexpression studies (Table 2). Identification of these factors and their induced molecular changes will be important in ameliorating reprogramming methods devising for instance virus-free iPS cells that are suitable for therapeutic use. In principle, regulatory factors that are able to replace the previously used set of reprogramming factors or improve the reprogramming efficiency have been described [77, 78]. For instance, Stat3 and Nanog are two pluripotency-related factors that are essential for the efficient progression to the pluripotent ground state [38, 49]. Here, we focus on identified factors related to their effect at defined stages of the reprogramming process and taking into account our proposed gene network motif behavior of the reprogramming process. For instance, fast gene expression in response to accelerating factors can lead to a faster response through the gene network cascade, leading to more efficient reprogramming (Fig. 3).

Most of the reprogramming accelerators that involve modifications of the epigenetic composition have been noted to display their effect in the late phase of reprogramming typically providing gradual changes in the epigenetic profile. In this context, de novo methyltransferases have been shown to be dispensable for reprogramming into an initial pluripotent state, but to be essential for establishing fully reprogrammed stem cells with a developmental potential similar to ESCs [32]. Moreover, 5-aza-2′-cytidine, a DNA methyltransferase inhibitor, has been shown to enhance reprogramming efficiency and accelerate the reprogramming dynamics only after 8 days [23]. Also, activation-induced cytidine deaminase, inducing general promoter demethylation, allows to initiate reprogramming toward pluripotency [14]. Wdr5, a mediator of H3K4 methylation, has been shown to act as a regulator of ESC self-renewal and to play a role during the early reprogramming phase [31]. Moreover, inhibition of histone deacetylase inducing a global increase in histone and non-histone protein acetylation thereby massively reversing repressive or activating epigenetic traits results into a more effective iPS cell generation [34]. More specifically, reduced expression of the Polycomb group protein complex 2 (PRC2) component genes typically causing defective trimethylation at histone H3K27 is positively correlated with better developmental genes and death of somatic cloned embryos, indicating that PRC2-induced H3K27me3 levels are essential for effective reprogramming [35]. Interestingly, binding of Oct4 to the Myod transactivation domain is thought to facilitate the generation of more accessible, permissive chromatin thereby enhancing the ability to continue with the reprogramming process [36]. Also, chromatin remodeling components such as the ATP-dependent BAF chromatin remodeling complex (BAF refers to mammalian Switch/Sucrose NonFermentable like Brg1/Brm associated factors) are described to have a clear increase in reprogramming efficiency [33]. In addition to regulatory factors exhibiting effects on the epigenetic composition, factors that act as signaling molecules have been shown to directly improve reprogramming efficiency and dynamics. For example, BMP-dependent induction of two microRNAs...
families, that is, miR-205 and miR-200, has been linked to induction of the early MET transition phase and to accelerate reprogramming progression into the late phase [24]. Furthermore, stimulation of the Wnt signaling and/or Notch pathways, involving cell proliferation, survival, apoptosis, and differentiation, have been shown to increase reprogramming efficiency [37]. Also inhibition of tumor suppressor genes such as Ink4A/Arf and P53 has been shown to improve the reprogramming efficiency, the latter has been shown to act in a proliferation-dependent manner, providing increasing efficiency only by allowing cells to divide faster [29, 30].

In conclusion, additional factors that can be classified as regulators of the epigenetic composition, tumor suppressors, or regulators of signaling cascades have been noted to affect the reprogramming efficiency and dynamics acting in a defined fashion. To verify and test mechanistic behavior of the distinct phases of the reprogramming process, it will be important to use such additional regulatory factors. Epigenetic modifiers might be helpful to determine the role of for example de novo methyltransferases in the reprogramming process (for review [79]). Inhibition of the regulatory factors acting in the tumor suppressor p53-p21 pathway could be useful to study the effects of cell death and proliferation during reprogramming [80], while signaling molecules could aid to toggle crucial cellular switches such as the MET transition during reprogramming. Transcription factors such as Stat3 and Nanog, which are essential for the progression to a fully pluripotent ground state, will be interesting subjects to determine the rate-limiting characteristics of the AND gate motifs discussed in Figure 2.

OUTLOOK

Therapeutic use of iPS cells is still not feasible due to the variability of reprogramming, the existence of partially reprogrammed cells, and the relative slowness of the process [17]. In order to tackle these challenges, a comprehensive molecular and systems level understanding of induced pluripotency is essential. In this review, we have provided a timeline of the molecular events (Table 1; Fig. 1) and gene network motifs (Figs. 2, 3) to capture the dynamic behavior of the reprogramming process. The gene networks and timeline provided here are of course a simplification of reality, but they aid in describing and predicting systems behavior of induced pluripotency. Knowledge of the distinct phases that make up reprogramming is essential to for example toggle and study specific reprogramming steps or alter reprogramming dynamics by changing the epigenetic composition thereby interfering with the permissiveness of chromatin (Table 2). Systems level understanding of induced pluripotency is important to identify and test factors that can potentially accelerate reprogramming (Table 2) for instance interfering with the proposed gene network motifs (Figs. 2, 3). This approach could be very helpful in removing the barriers that still hinder the therapeutic use of iPS cells.

REFERENCES


www.StemCells.com

ACKNOWLEDGMENTS

We thank burobrouns.nl for creating the cartoon-like representation of the reprogramming dynamics and advice on illustrating the ideas expressed in this review. This work was supported in part by grants from the Netherlands Organization for Scientific Research (ALW-NWO-Meevord 836.07.001 to P.J.V.).

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors declare no potential conflict of interest regarding financial authorship or publication of this article.