Case Essay

Synthetic gene network for the efficient generation of induced pluripotent stem cells from adult somatic cells

Nicolas Jaccard*

Supervisors: Prof. John Ward** | Prof. Alexey Zaikin***

Center for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London*
Department of Structural and Molecular Biology, University College London, London**
Department of Mathematics, University College London, London***

Embryonic stem cells derived from the inner cell-mass of the blastocyst are self-renewable, meaning they can undergo an infinite number of divisions. They are also pluripotent and can give rise to most adult cell types through differentiation, a genetically regulated process during which cells gain specialized phenotypes. These unique properties make them a potentially unlimited source of material for developmental biology, drug discovery and regenerative medicine. Recently, it was shown that it is possible to generate pluripotent cells from adult somatic cells by reprogramming them using four transcription factors: Oct4, c-Myc, Sox2 and Klf4. This approach alleviates the concerns related to the embryonic origins of the cells and promises patient-specific therapies. However, current reprogramming protocols remain very inefficient, mainly due to a lack of control of the transcription factors expression. Synthetic biology, by combining knowledge from various disciplines including mathematics, molecular biology and engineering, provides a framework for the rational design of complex genetic networks that can be used to improve existing characteristics of a cell or create *de-novo* functionalities, such as advanced gene expression regulatory mechanisms.

This case essay examines the use of a synthetic regulatory cascade to improve the reprogramming process of somatic cells by tightly controlling the expression of the four transcription factors. The design is based around three compatible antibiotic responsive expression systems: TET_{OFF}, E_{OFF} and PIP_{OFF}, repressed by tetracycline, erythromycin and pristinamycin respectively. Numerical simulations based on a system of differential equations have shown that it is indeed possible to generate complex expression patterns over the typical induction period, allowing for a degree of control not achievable with existing approaches. An evaluation protocol is also proposed to experimentally compare the performances of the regulatory cascade against traditional reprogramming methods.

Precise temporal expression of known amounts of transcription factors might be used as a tool to better understand the molecular mechanisms involved in the reprogramming process and, as a directed consequence, might lead to the establishment of robust and highly efficient induction protocols that would allow for the generation of the large quantity of cells that will most likely be required for future research and therapeutic applications.

1. Introduction

This essay summarizes the current state of synthetic biology in mammalian cells and the latest developments in somatic cell reprogramming before exploring the use of a hypothetical synthetic regulatory cascade for the generation of induced pluripotent stem cells.

1.1 Synthetic gene networks in mammalian systems

Since the discovery of the DNA structure and the subsequent cracking of the genetic code, genes and their functions have been extensively studied. Forward genetics (also referred as genetic screening), pioneered by Nusslein-Volhard and Wieschaus[1], was first used to determine the genetic basis of a given phenotype. The inverse problem was also studied with reverse genetics, method that consists in altering the transcription and/or translation mechanisms of specific genes and detecting the resulting effects on the traits of the studied organism. However, the findings of the human genome project[2] showed that a piece of the puzzle was missing: about 30'000 protein-coding transcripts were identified, which was a surprisingly low number considering a worm (*Caenorhabditis elegans*) was found to possess about 20'000 genes[3]. It was then clear that biological complexity could not solely be a function of the number of genes but must also involve other regulatory mechanisms. In an effort to understand how elements identified using traditional approach interact and form networks, a holistic and interdisciplinary strategy was adopted by combining different techniques borrowed from engineering sciences such as high throughput analysis and computational modeling, leading to the creation of a new field, systems biology.

As the understanding of fundamental mechanisms increases, so does the ability to engineer DNA. Thanks to the fact the genetic code is mostly universal[4], expression of a set of exogenous genes in a cellular host give rise to new traits borrowed from other organisms, as it is the case with cells turning green under UV light when transfected with the GFP (green fluorescent protein) coding genes isolated from a jellyfish[5]. Advanced molecular biology cloning techniques allow to routinely assemble genetic elements from diverse sources to confer cells a specific trait, such as the capacity to produce recombinant human insulin[6]. Whereas this genetic engineering approach solely relies on pre-existing biological parts and is limited to simple systems, synthetic biology provides a framework combining knowledge from molecular biology, engineering and mathematics to rationally design biological networks (such as gene circuits or metabolic pathways), molecules (e.g. proteins, peptides) and, in the not so distant future, whole organisms.

Many complex synthetic circuits have been built using prokaryotic systems, including oscillators[7], pulse generators[8], time-delays[9] or signal cascades[10]. *Vibrio fisheri* quorum-sensing circuit was re-engineered and expressed in an *E. coli* chassis (cellular host) to promote multicellular patterns formations[11]. A prey-predator artificial ecosystem based on cell-cell communication was created and showed good correlations with mathematical population dynamics models[12]. Advanced sensing and decision-making based on multiple input signals were integrated to allow *E. coli* cells to detect, invade and kill cancer cells[13]. Such examples show the great promises of synthetic biology but remain mostly limited to prokaryotic chassis that cannot be used in a clinical context. The next major milestone will be the design of complex synthetic circuits that can function in mammalian cells and be used as part of novel therapeutic strategies.

While the basic elements involved in mammalian cells gene expression are similar to what is found in prokaryotic systems, their interactions and underlying mechanisms are much more complex. For instance, the expression of a transgene can vary dramatically depending on its location in the host's genome. Chromatin can be of two forms: heterochromatin and euchromatin, describing condensed and extended DNA regions respectively. For the transcription to be initiated, it is important that both transcription factors and polymerase have access to the gene and its regulatory sequences. Euchromatin, by being more relaxed, will more likely be actively transcribed whereas genes in heterochromatin might end up silenced. Chromatin conformal changes are mainly driven by chromatin remodeling proteins such as histone deacetylase. DNA methylation also plays a major role in epigenetic regulation of gene expression, either by helping the recruitment of aforementioned

remodeling proteins or by blocking access to key regulatory regions of a gene[14]. The use of specific DNA elements such as matrix-attachment regions or insulators protects the transgene from inactivation and help achieving robust expression in different clones despite having different integration locations. Another difference comes from the fact that transcription and translation occur in two separate compartments, resulting in an additional lag time between the two processes due to molecular transport (which is mostly negligible in prokaryotes) implying that motifs that rely on coupled transcription and translation regulation processes might be more difficult to engineer in mammalian cells.

Despite the multiple hurdles related to mammalian chassis, most basic components that are required to build complex synthetic gene networks are now available in such systems. Inducible promoters are used as the main control elements in these circuits, most of them being based on bacterial antibiotic response regulators (Figure 1). They have low interference with endogenous activities of mammalian cells and multiple systems can be combined in a single synthetic network[15].

Kramer *et al*[16] combined several of these promoters to create a regulatory cascade capable of modulating the level of expression of a gene in response to the type of inducer present in the culture medium. Similarly, using combinations of inducible promoters, a set of logic gates (NOTIF, NAND, OR, NOR and an inverter) were demonstrated in mammalian chassis[17], enabling the processing of multiple inputs. An alternative way to achieve Boolean decision-making made use of RNA interference and allowed the evaluation of up to five logic variables[18].

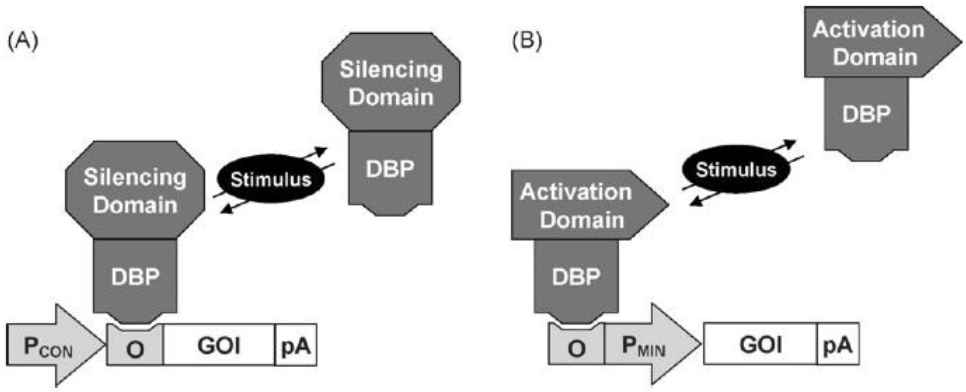

Figure 1 (**A**) Repressor system: the operator (repressor site) is placed downstream a constitutive (unregulated) promoter. Transcription can be blocked by steric hindrance or the DNA binding protein (DBP) can be fused with a trans-silencing domain (usually Kruppel-associated box protein or KRAB). (**B**) Activation system: the operator (activator site) is placed upstream a minimally active promoter. The DNA binding protein (DBP) is fused with a transactivating domain (usually VP16 of *Herpes simplex*) and will initiate the transcription upon binding. Both systems can be inducible with various stimuli. (Figure source: Greber and Fussenegger[19])

The presence of an inducer over a long time period might not be desired for some processes, especially when said inducer is an antibiotic. Although some basal leaking and low maximal expression levels were reported, a toggle (bistable) switch (allowing to switch between an OFF and an ON state) has been constructed using a mutually inhibitory arrangement[20], showing that it is indeed possible to produce a long-term response to a transient signal. As an alternative, a switch integrating current and historic input signals (hysteretic switch) has also been demonstrated in mammalian cells[21].

These modules (control systems, input processing and switches) represent the basis for the construction of sophisticated synthetic network for mammalian chassis: Weber *et al.* built a time-delay circuit[22] and Tigges *et al.* put together an extremely well-made design for a genetic oscillator[23]. Not only these two examples show that mammalian chassis can now be the host of synthetic gene networks that achieve a degree of complexity similar to their prokaryotic counterpart, but also that they can produce models that might help understand fundamental biological mechanisms (biological clocks in this case).

As synthetic biology is inspired by fields such as engineering, there is a need for formalism and standardization so knowledge can be shared and re-used. Two main standards are used in the community to exchange mathematical models: the systems biology markup language[24] (SBML) and the cellular markup language[25] (CellML). In addition to the storage of mathematical expressions, they allow the annotation of parts and their documentation. They are free to implement and many tools are available for the creation, simulation and validation of synthetic biology models[26].

1.2 Induced pluripotent stem cells

Embryonic stem cells are found in the inner cell mass of the blastocyst of a developing embryo[27]. They can differentiate into any cell type of the three germ layers (endoderm, mesoderm and ectoderm)[28] and can self-renew, meaning they can divide and proliferate as long as specific signals are present[27]. Permanent mouse embryonic stem cells lines[29] have been extensively used as models in developmental biology, leading to major advances in that field[30]. After the derivation of the first human embryonic stem cell lines (hESC)[27], research has been conducted to develop therapeutic strategies involving the *in-vitro* generation of tissues that could then be transplanted in order to repair damaged organs. These therapies might provide a treatment for neurodegenerative diseases[31] or for brain damages after a stroke[32]. Embryonic stem cells can also be a source of biological material for research purposes, including drug discovery[33] or toxicology screenings[34].

Recent studies have shown that, in contrary to what was believed, the differentiation process is not irreversible. Takahashi *et al.* were able to reprogram adult mouse fibroblast cells back to a pluripotent state[35]. These so-called induced pluripotent stem cells (iPS) could help alleviate some of the fear regarding the ethics of using cells originating from a human embryo. Another advantage is that the cells would be patient-specific and thus greatly reduce the risk of immunoresponse upon transplantation.

Surprisingly, the reprogramming of cells only required the expression of four transcription factors: Oct4, Sox2, c-Myc and Klf4. In a matter of months, it was shown that the same factors could be used to reprogram human somatic cells into pluripotent cells that were very similar to embryonic stem cells, were able to differentiate into cells from the three germ layers and were positive to all pluripotency markers[36]. This discovery generated an unprecedented scientific interest. Studies are already showing that reprogrammed cells hold great promises in a clinical context: mice were rescued from all sickle cell anemia symptoms using hematopoietic precursors obtained from autologous (from the individual) iPS cells[37]. It has also been demonstrated that neurons can be generated from rat fibroblast cells and their subsequent transplantation greatly improved symptoms of individuals with Parkinson's disease[38]. Functional human insulin secreting cells and cardiomyocytes[39] were successfully obtained from somatic cells, paving the way for future regenerative medicine therapeutic strategies[40].

If these promises are to be fulfilled, many obstacles need to be overcome. The genes coding for the reprogramming factors were transduced into cells using lentiviral vectors (viral particles part of the retrovirus family) because of their high efficiency and their ability to integrate the genome of quiescscent cells[41]. This seemingly random integration can lead to the activation of oncongenes (or the inactivation of tumor repressor genes) via a process called insertional mutagenesis[42]. The fact that all four genes were transduced separately increased that risk. It is possible to limit the chance of such an event to occur by cloning all the genes coding for the transcription factors on a single polycistronic vector. Using loxP/Cre technology (controlled site-specific recombination events), the exogenous DNA introduced by transduction can be excised from the host genome, thus producing transgene-free cells that are thought to be safer for clinical uses[43]. Another solution would be the use of integration-deficient lentiviral vectors, which do not integrate in the host genome but remain as circular vectors episomes while retaining their high transduction efficiency[42]. Similarly, virus-free induction was possible using a non-viral plasmid (that can either be transiently expressed or integrated at a specific site in the genome if it contains the *piggyBac* transposon elements)[44]. Recently, Zhou *et al.* used recombinant transcription factors that were fused to a poly-arginine transduction domain to induce pluripotent stem cells without relying on the expression of exonous genes[45]. This method could be the solution to most of the concerns regarding the safety of

induced cells and their reprogramming protocols, although the purification of such proteins remains cost prohibitive. The different transduction approaches are summarized in Figure 2.

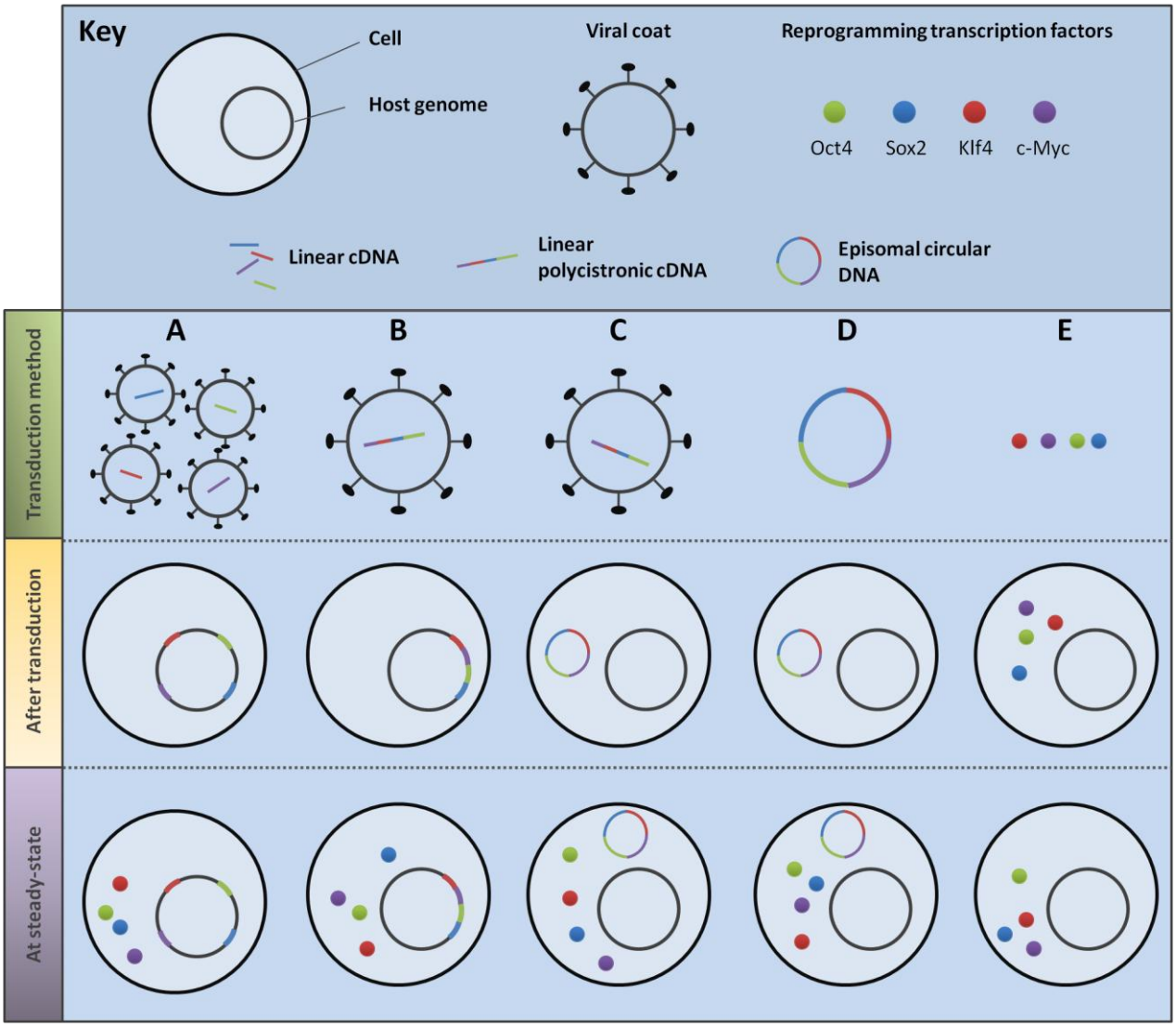

Figure 2 Different transduction approaches to reprogram somatic cells into induced pluripotent stem cells. The first step shows the transduction material, the second is the state of the cell moments after the transduction and the third represents long-term (steady-state) behavior (**A**) Four different viral vectors containing cDNA for each transcription factors, leading to integration at four different locations in the host genome. (**B**) Viral vector containing a polycistronic cDNA of all four transcription factors leading to integration at a single site in the host genome. (**C**) Integration-deficient viral vector containing a polycistronic cDNA of all four transcription factors leading to a transient expression without integration (**D**) Circular DNA coding for all four transcription factors, leading to a transient expression without integration (**E**) Transduction of purified transcription factors

Great efforts have been put forward understanding how these transcription factors (Oct4, Sox2, Klf4 and c-Myc) are able to revert the cell to a pluripotent state. Oct4 is expressed in the inner cell mass at early stages of the development but is downregulated in somatic cells, playing a major role in pluripotency regulation[46]. Small variations from its optimal expression level can lead to differentiation[47]. In contrast, Sox2 is also present in some precursor cells and is actually required for the viability of adult neuronal cells[48]. Klf4 is expressed during growth arrest and is downregulated in actively dividing cells, hinting at his possible role in DNA synthesis and cell cycle regulation[49]. The last reprogramming factor is c-Myc, a multidomain transcription factor involved in the regulation of more than 15% of the human genome and implicated in many key cellular mechanisms such as growth, differentiation or proliferation. It is upregulated in more than 10% of human cancers[50]. Other transcriptions factors were also linked with various forms of cancer and oncogenic abilities of cells[51-53]. Another key regulator, which is not part of the reprogramming cocktail as its exogenous expression doesn't seem to have

any effect on the cell state, is the transcription factor Nanog: its presence is required to maintain cultured embryonic stem cells in a pluripotent state[53], mostly by repressing key genes involved in the differentiation process[54]. While the exact molecular events leading to a reprogrammed state are still mostly unknown, a two-stage mechanism has been proposed by some studies (Figure 3).

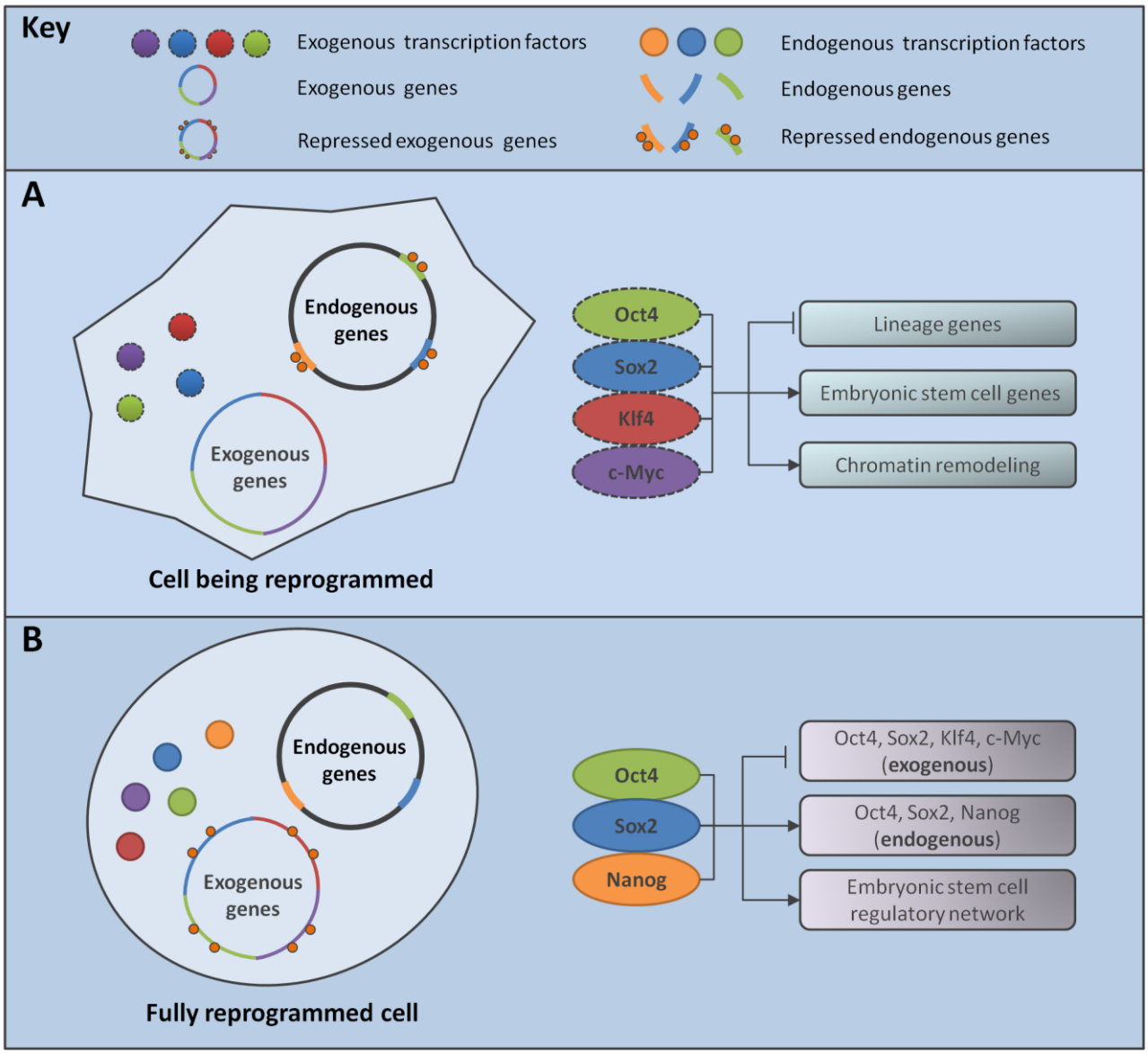

Figure 3 A hypothetic molecular mechanism for the reprogramming of somatic cells (**A**) The cocktail of reprogramming transcription factors is transduced into the cell. Their first effect seems to be the repression of lineage genes and the activation of embryonic stem cell genes. The chromatin is remodeled to allow for the activation of endogenous Oct4, Sox2 and Nanog. (**B**) Oct4, Sox2 and Nanog are activated through chromatin remodeling, leading to the repression of exogenous genes and a further activation of their own expression (auto feedback loop). The regulatory network of embryonic stem cells is re-established, leading to the complete reprogramming of the cell. (Adapted from Scheper *et al*[55], Rodda *et al*[56], Wilmut[57] and Varas *et al*[58])

2. A synthetic gene network for efficient somatic cells reprogramming

The reprogramming of somatic cells is a very inefficient process: about 0.1% of the cells are selected and confirmed to be pluripotent. It is believed that one of the main cause might be inappropriate expression levels of the transcription factors[55]: small variations of Oct4 lead to the differentiation of the cells and a precise balance between Klf4 and c-Myc has to be maintained through the process. Existing methods employ an all or nothing induction scheme that does not allow fine tuning individual gene expression. This case essay examines the use of a novel reprogramming approach using a synthetic regulatory cascade that allows temporally controlling the expression levels of the reprogramming transcription factors during the induction process.

2.1 Synthetic gene network design

Regulatory cascades are a recurring motif in biological systems. As mentioned in the introduction, a synthetic cascade has been implemented in a mammalian chassis by Kramer *et al*[16] using a combination of compatible antibiotics responsive expression systems (see Table 1). They are composed of a promoter, an inducing molecule and a repressing antibiotic (Tetracycline, Erythromicin or Pristinamycin) that has been licensed for clinical uses. Results have shown that it is indeed possible to have the expression of the gene of interest modulated by the presence of the different repressors in the culture medium. For this work, the described design has been adopted and extended with additional IRES (internal ribosome entry sites) sequences, allowing for the regulation of additional genes.

Table 1 Description of the TET_{OFF}, E_{OFF} and PIP_{OFF} antibiotic responsive expression systems (Adapted from Kramer *et al.*[16])

System name	Promoter	Inducer	Repressor
TET_{OFF}	$P_{hCMV*-1}$	tTA	Tetracycline
E_{OFF}	P_{ETR2}	ET1	Erythromicin
PIP_{OFF}	P_{PIR8}	PIT	Pristinamycin

The proposed construct combine all three systems (Figure 4) to control the expression of the four reprogramming factors (Oct4, Sox2, c-Myc and Klf4), which are named factor 1 to 4 for simulation purposes. Although antibiotics are used to shutdown parts of the network, leaking (basal) expression can lead to slow accumulation of activators that keep downstream elements of the network active. For instance, the addition of tetracycline will drastically reduce the expression of 'Factor1' but the basal expression of ET1 might be sufficient to activate P_{ETR2}. These scenarios are shown in the numerical simulation section.

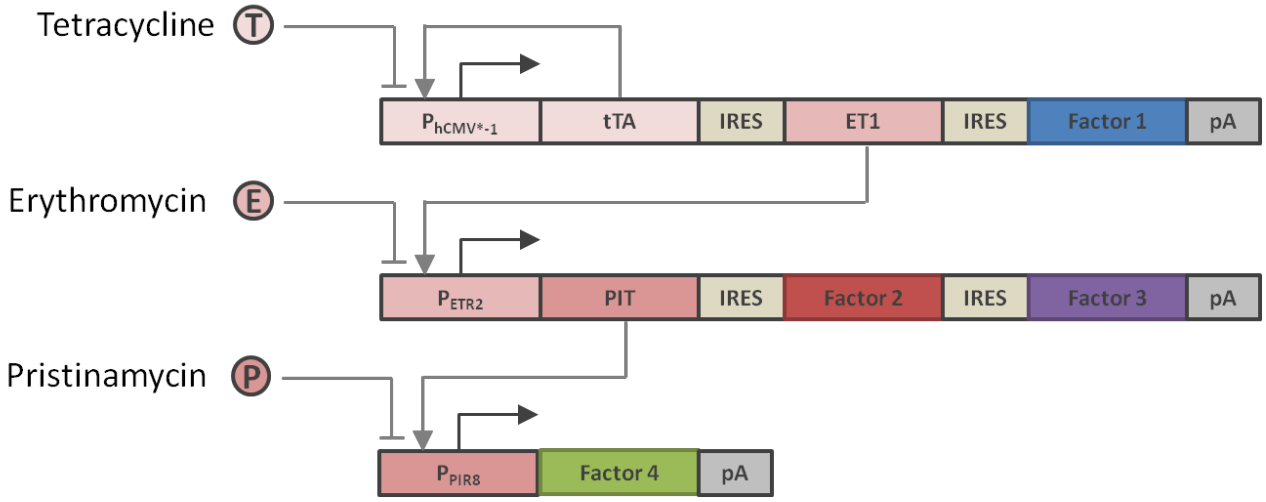

Figure 4 Design of the synthetic gene network for somatic cell reprogramming. $P_{hCMV*-1}$ (tetracycline- responsive promoter) controls the expression of tTA (tetracycline-dependant transactivator), which is the autoregulator of the promoter. It also controls the expression of ET1 (macrolide-dependant transactivator) and one of the reprogramming factor. ET1 activates P_{ETR2} (macrolide-responsive promoter). This activation leads to the expression of two of the factors in addition to PIT (streptoramin-dependant transactivator) which activates P_{PIR8} (streptogramin-responsive promoter), driving the expression of the fourth reprogramming factor. Each of the promoters can be repressed using the corresponding antibiotic (Tetracycline, erythromycin, pristinamycin)

2.2 Numerical simulation

Different approaches have been used to mathematically describe gene networks, including boolean models[59], differential equation models[60] or hybrid models[61]. In this work, a differential equation model was used:

$$\frac{dY}{dt} = f(X) + b - \gamma Y \qquad \text{(Equation 1)}$$

Where Y [-] is the product of the regulated gene, f(X) is a function describing the effect of regulator X [-] (usually a hill function), b [min^{-1}] the basal expression of Y and γ [min^{-1}] the decay rate of Y. As proposed by Kramer *et al.*[16], f(X) can be of the following form in case of a positive regulation of Y by X:

$$\frac{dY}{dt} = \frac{X \cdot \alpha}{K + X} + b - \gamma Y \qquad \text{(Equation 2)}$$

Where K is the activation constant [-] and α [min^{-1}] the effective rate of synthesis of Y, a lumped parameter describing the net effect of polymerase binding, open-complex formation, transcript elongation, transcript termination, ribosome binding and polypeptide elongation[62]. It was decided to use this dimensionless model due to its success at predicting synthetic regulatory cascades behavior in mammalian cells[16].

The following system was used to describe the synthetic regulatory cascade (Figure 4):

TET_{OFF} system with an auto-induction from tTa, controlling ET1 and F1 (Factor 1) expressions:

$$\frac{dtTa}{dt} = \frac{tTa \cdot \alpha_1}{K_1 + tTa} + b_1 - \gamma_{tTa} \cdot tTa \qquad \text{(Equation 3)}$$

$$\frac{dET1}{dt} = \frac{tTa \cdot \alpha_1}{K_1 + tTa} + b_1 - \gamma_{ET1} \cdot ET1 \qquad \text{(Equation 4)}$$

$$\frac{dF1}{dt} = \frac{tTa \cdot \alpha_1}{K_1 + tTa} + b_1 - \gamma_{F1} \cdot F1 \qquad \text{(Equation 5)}$$

E_{OFF} system induced by ET1, controlling PIT, F2 and F3 (Factor 2 and 3) expressions:

$$\frac{dPIT}{dt} = \frac{ET1 \cdot \alpha_2}{K_2 + ET1} + b_2 - \gamma_{PIT} \cdot PIT \qquad \text{(Equation 6)}$$

$$\frac{dF2}{dt} = \frac{ET1 \cdot \alpha_2}{K_2 + ET1} + b_2 - \gamma_{F2} \cdot F2 \qquad \text{(Equation 7)}$$

$$\frac{dF3}{dt} = \frac{ET1 \cdot \alpha_2}{K_2 + ET1} + b_2 - \gamma_{F3} \cdot F3 \qquad \text{(Equation 8)}$$

PIP_{OFF} system induced by PIT, controlling F4 (Factor 4) expressions:

$$\frac{dF4}{dt} = \frac{PIT \cdot \alpha_3}{K_3 + PIT} + b_3 - \gamma_{F4} \cdot F4 \qquad \text{(Equation 9)}$$

The models parameters were chosen according to information found in the literature. The constants of activation (K_1, K_2 and K_3) were set to 1000 (corresponding to half–maximal expression levels[16]). The effective rates of synthesis (α_i) correspond to the values obtained for a biphasic switch in prokaryotic cells[62]: in absence of the

corresponding antibiotic, α_i is equal to 10 min^{-1} in all cases, meaning the strengths of the promoters are considered equal (as determined experimentally[63]). As the presence of antibiotics has been shown to completely inhibit the action of the activators for these systems[63], α_i was set to 0 min^{-1} when they are added to the medium. Following from previous hypotheses, the basal expressions were also supposed to be the same for all three promoters ($b_1=b_2=b_3$) and were set to 0.5 min^{-1} [16]. The degradation rates of the different species were set to 0.0005 min^{-1} (as measured for heterologous GFP in mammalian cells)[64].

The initial amount of activators and factors were set to zero. The cascade is initialized by the basal expression of the TET_{OFF} system (Equation 3). It was decided to run the simulation for an equivalent of 18 days, which corresponds to a typical reprogramming time using similar approaches[44]. The system of differential equations was solved numerically using Mathematica (Version 7.0.1.0, Wolfram Research Inc.) and data was exported to Excel (Version 2007 SP2, Microsoft Corporation).

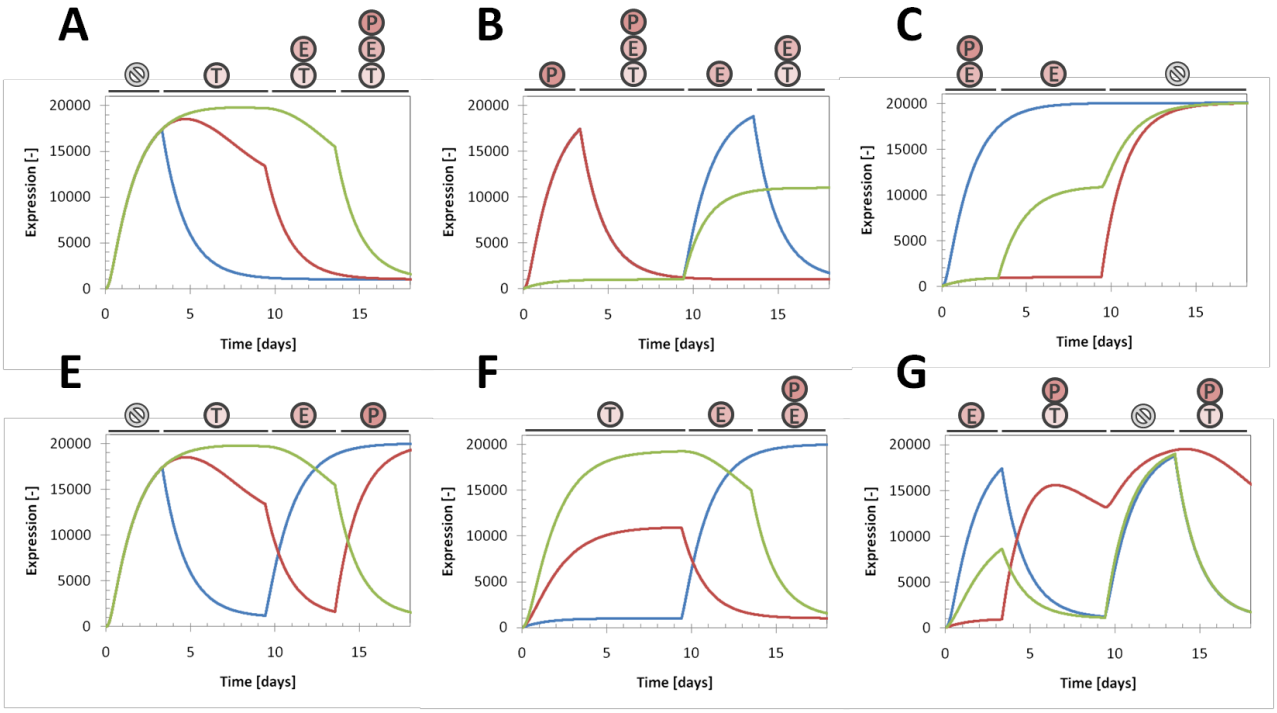

Figure 5 Numerical simulations of the synthetic cascade. The curves represent the expression of the different reprogramming factors: Factor 1 in blue (—), Factor 2 and 3 in red (—), Factor 4 in green (—). The symbols above the graphs indicate the periods of application of specific antibiotics (T =Tetracycline, E =Erythromycin, P =Pristinamycin, ⊘ = no antibiotic applied)

The resulting graphs (Figure 5) show only three curves as two of the factors (Factors 2 and 3 on Figure 4) are considered to have similar expression patterns due to the fact they are under the control of the same promoter. Vastly different profiles emerge depending of the antibiotic application strategy. All factors can have similar expression levels early on and then be switched off at different points in time (A on Figure 5). On the contrary, they can be turned on sequentially (C on Figure 5). Other intermediate expression patterns are also observable, such as oscillations (E and G on Figure 5). It confirms that the expression of a factor is most influenced by the repressor directly inhibiting its cognate promoter while the repression of upstream promoters have less of an effect due to the basal expression keeping downstream parts of the circuit active.

These results seem to show that complex expression profiles can, at least *in-silico*, be generated using the proposed synthetic regulatory cascade. However, as it is, the model is probably not a good representation of the biological reality, mainly due to the way parameters were chosen. Effective rates of synthesis were determined in prokaryotic systems, which mean they are probably not suitable for mammalian hosts. They were also chosen to be the same for all the genes but they must surely depend on specific features such as the length of the gene coding sequence or the structural complexity of its product. Another oversight might concern the rates of decay

that were determined using green fluorescent protein, which might not be representative of other proteins in term of half-life. Another unknown concerns the synthesis rates for genes after an IRES sequence. Some evidences tend to show that, in that case, initiation is twice as slow compared to regular cap-dependant initiation[65]. Experiments should be performed to determine conditions-specific parameters that could in-turn be used to refine the model. The next step could be the use of advanced analysis packages such as COPASI (COmplex PAthway SImulator): instead of relying on a parameter describing all the critical steps (the effective rate of synthesis α), these could be individually modeled and further analyzed.

2.3 Performance evaluation protocol

While the model is useful to get a sense of what is achievable with such a regulatory cascade, it is not possible to predict which expression pattern will translate into the best reprogramming efficiency. Here are the essential steps of an evaluation protocol (Figure 6) that could be used to experimentally determine which antibiotic application strategy is the most optimal for somatic cell reprogramming using the proposed gene cascade and to compare its performance against existing methods.

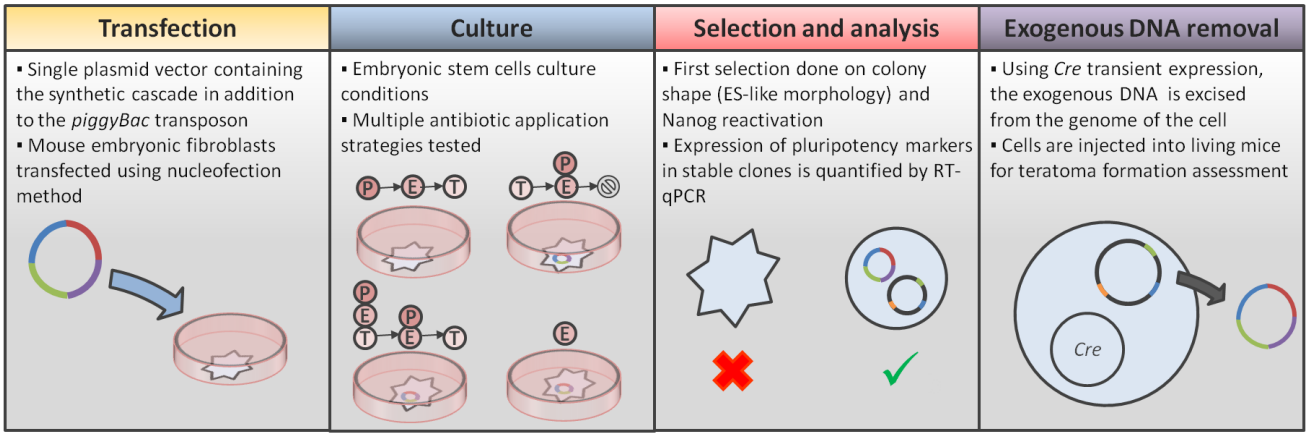

Figure 6 Evaluation protocol to determine the best antibiotic application strategy for optimal reprogramming results using the proposed synthetic gene cascade

Mouse embryonic fibroblast (MEF) cells are often used as models for reprogramming experiments[44]. The chosen vector is a plasmid harboring the genetic cascade elements (Figure 4, with Factors representing Oct4, Sox2, c-Myc and Klf4) as well as the *piggyBac* transposon sequences that allow site-specific integration into the host genome and subsequent excision when *Cre* is expressed by the cell[66]. The plasmid can either be constructed using molecular biology cloning techniques (as all parts are readily available) or be synthesized[67]. Concerning the actual transfection, nucleofection proved to be an efficient method for the introduction of exogenous DNA into MEF cells[44]. Cells are then cultured in embryonic stem cells conditions (culture medium containing differentiation repressors such as leukemia inhibitor factor[68] and with growth surfaces coated in gelatin[69]) for a duration of 18 days during which different antibiotic application strategies are tested on individual cultures. To determine which cells have both received the transgene and started reprogramming, the first selection is done on colony shapes (negative cells retain typical fibroblast morphology) and on Nanog reactivation (an early sign of successful reprogramming, see induced pluripotent stem cells introduction). Positive cells are then re-plated, again in embryonic stem cells conditions. In order to obtain transgene-free clones, cells are transfected with a plasmid containing the *Cre* sequence, triggering the excision of the exogenous DNA from the host genome. The stability of the pluripotency state can be evaluated by determining the expression level of specific markers (Oct4, Sox2 and Nanog) using real-time quantitative PCR (RT-qPCR). The ability of the cells to form teratomas (encapsulated tumor containing tissues from all three germ layers) upon injection into a living mouse can also be evaluated as a pluripotency marker[70].

2.4 Conclusion and outlooks

A novel somatic cells reprogramming strategy has been examined. A synthetic regulatory cascade design based on antibiotic response-systems has been extended so it can be used to control the expression of the four factors required for the induction of pluripotent stem cells. Numerical simulations based on a system of differential equations showed that various complex expression patterns are achievable, allowing a degree of control that is not obtainable with existing methods. Being able to have specific levels and intervals of expression for the four factors allows studying their individual effects during the reprogramming process and thus will shed some light on the underlying molecular mechanisms. The direct consequence might be the establishment of robust and highly efficient protocols based on a precise understanding of the process whereas current approaches rely on bulk expression of the four factors and result in a small fraction of the cells getting completely reprogrammed (Figure 7).

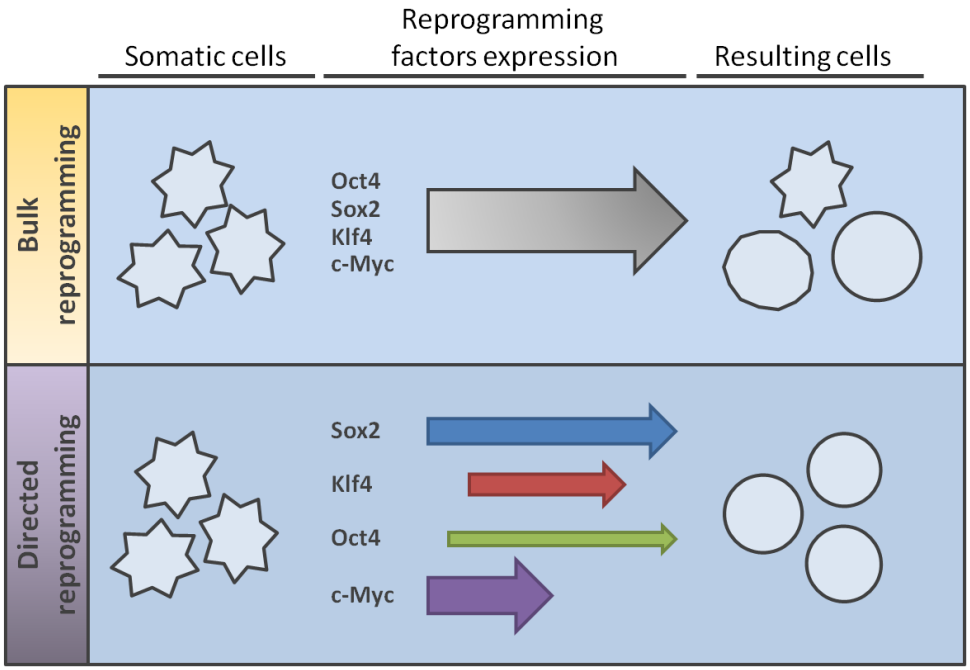

Figure 7 Comparison between the traditional reprogramming methods that use bulk expression of the four transcription factors and the proposed strategy that allows to have specific expression levels and intervals for each of the factor. The bulk reprogramming is a very inefficient process leading to a heterogenous population while directed reprogramming, through a precise understanding of the underlying molecular mechanisms, would result in a homogenous population of induced pluripotent stem cells.

Such results would represent the required first step for a transition from the laboratory to the clinic. However, some aspects of the approach can be further improved. The proposed design uses antibiotics as effecting molecules to control the expression levels of the different reprogramming factors, choice dictated by the fact the necessary biological parts were readily available and characterized. Due to heavy regulations concerning the use of antibiotics in clinical contexts, it might be necessary to develop an equivalent system using other types of inducible promoters, such as the recently described cumate-based control system[71]. The addition of other compatible inducible systems would also allow having four individually controlled factors as opposed to three with the current design (two of the genes being under the control of the same promoter). It would also be necessary to devise strategies to control the amount of DNA delivered to ensure homogenous expression profiles among the transfected population.

Despite some limitations, this novel approach to cell reprogramming could potentially lead to efficient protocols allowing for the generation of the large quantities of cells that are most likely going to be required for both research and therapeutic applications.

3. References

1. Nusslein-Volhard, C. and E. Wieschaus, *Mutations affecting segment number and polarity in Drosophila.* Nature, 1980. **287**(5785): p. 795-801.
2. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
3. *Genome sequence of the nematode C. elegans: a platform for investigating biology.* Science, 1998. **282**(5396): p. 2012-8.
4. Koonin, E.V. and A.S. Novozhilov, *Origin and evolution of the genetic code: the universal enigma.* IUBMB Life, 2009. **61**(2): p. 99-111.
5. Prasher, D.C., et al., *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene, 1992. **111**(2): p. 229-33.
6. Nossal, G.J., *Human insulin through recombinant DNA technology.* Med J Aust, 1980. **2**(6): p. 295-6.
7. Fung, E., et al., *A synthetic gene-metabolic oscillator.* Nature, 2005. **435**(7038): p. 118-22.
8. Basu, S., et al., *Spatiotemporal control of gene expression with pulse-generating networks.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6355-60.
9. Mangan, S., A. Zaslaver, and U. Alon, *The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks.* J Mol Biol, 2003. **334**(2): p. 197-204.
10. Rosenfeld, N., et al., *Gene regulation at the single-cell level.* Science, 2005. **307**(5717): p. 1962-5.
11. Basu, S., et al., *A synthetic multicellular system for programmed pattern formation.* Nature, 2005. **434**(7037): p. 1130-4.
12. Balagadde, F.K., et al., *A synthetic Escherichia coli predator-prey ecosystem.* Mol Syst Biol, 2008. **4**: p. 187.
13. Anderson, J.C., et al., *Environmentally controlled invasion of cancer cells by engineered bacteria.* J Mol Biol, 2006. **355**(4): p. 619-27.
14. Nakao, M., et al., *Regulation of transcription and chromatin by methyl-CpG binding protein MBD1.* Brain Dev, 2001. **23 Suppl 1**: p. S174-6.
15. Weber, W., et al., *Macrolide-based transgene control in mammalian cells and mice.* Nat Biotechnol, 2002. **20**(9): p. 901-7.
16. Kramer, B.P., W. Weber, and M. Fussenegger, *Artificial regulatory networks and cascades for discrete multilevel transgene control in mammalian cells.* Biotechnol Bioeng, 2003. **83**(7): p. 810-20.
17. Kramer, B.P., C. Fischer, and M. Fussenegger, *BioLogic gates enable logical transcription control in mammalian cells.* Biotechnol Bioeng, 2004. **87**(4): p. 478-84.
18. Rinaudo, K., et al., *A universal RNAi-based logic evaluator that operates in mammalian cells.* Nat Biotechnol, 2007. **25**(7): p. 795-801.
19. Greber, D. and M. Fussenegger, *Mammalian synthetic biology: engineering of sophisticated gene networks.* J Biotechnol, 2007. **130**(4): p. 329-45.
20. Kramer, B.P., et al., *An engineered epigenetic transgene switch in mammalian cells.* Nat Biotechnol, 2004. **22**(7): p. 867-70.
21. Kramer, B.P. and M. Fussenegger, *Hysteresis in a synthetic mammalian gene network.* Proc Natl Acad Sci U S A, 2005. **102**(27): p. 9517-22.
22. Weber, W., et al., *A synthetic time-delay circuit in mammalian cells and mice.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2643-8.
23. Tigges, M., et al., *A tunable synthetic mammalian oscillator.* Nature, 2009. **457**(7227): p. 309-12.
24. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, 2003. **19**(4): p. 524-31.
25. Lloyd, C.M., M.D. Halstead, and P.F. Nielsen, *CellML: its future, present and past.* Prog Biophys Mol Biol, 2004. **85**(2-3): p. 433-50.
26. Endler, L., et al., *Designing and encoding models for synthetic biology.* J R Soc Interface, 2009. **6 Suppl 4**: p. S405-17.
27. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
28. Odorico, J.S., D.S. Kaufman, and J.A. Thomson, *Multilineage differentiation from human embryonic stem cell lines.* Stem Cells, 2001. **19**(3): p. 193-204.
29. Hooper, M., et al., *HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells.* Nature, 1987. **326**(6110): p. 292-5.
30. Wobus, A.M. and K.R. Boheler, *Embryonic stem cells: prospects for developmental biology and cell therapy.* Physiol Rev, 2005. **85**(2): p. 635-78.
31. Goldman, S., *Stem and progenitor cell-based therapy of the human central nervous system.* Nat Biotechnol, 2005. **23**(7): p. 862-71.

32. Zhang, S.C., et al., *Human embryonic stem cells for brain repair?* Philos Trans R Soc Lond B Biol Sci, 2008. **363**(1489): p. 87-99.
33. Sartipy, P., et al., *The application of human embryonic stem cell technologies to drug discovery.* Drug Discov Today, 2007. **12**(17-18): p. 688-99.
34. Davila, J.C., et al., *Use and application of stem cells in toxicology.* Toxicol Sci, 2004. **79**(2): p. 214-23.
35. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
36. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
37. Hanna, J., et al., *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* Science, 2007. **318**(5858): p. 1920-3.
38. Wernig, M., et al., *Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5856-61.
39. Zhang, J., et al., *Functional cardiomyocytes derived from human induced pluripotent stem cells.* Circ Res, 2009. **104**(4): p. e30-41.
40. Tateishi, K., et al., *Generation of insulin-secreting islet-like clusters from human skin fibroblasts.* J Biol Chem, 2008. **283**(46): p. 31601-7.
41. Lever, A.M., P.M. Strappe, and J. Zhao, *Lentiviral vectors.* J Biomed Sci, 2004. **11**(4): p. 439-49.
42. Wanisch, K. and R.J. Yanez-Munoz, *Integration-deficient lentiviral vectors: a slow coming of age.* Mol Ther, 2009. **17**(8): p. 1316-32.
43. Soldner, F., et al., *Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors.* Cell, 2009. **136**(5): p. 964-77.
44. Kaji, K., et al., *Virus-free induction of pluripotency and subsequent excision of reprogramming factors.* Nature, 2009. **458**(7239): p. 771-5.
45. Zhou, H., et al., *Generation of induced pluripotent stem cells using recombinant proteins.* Cell Stem Cell, 2009. **4**(5): p. 381-4.
46. Pesce, M. and H.R. Scholer, *Oct-4: gatekeeper in the beginnings of mammalian development.* Stem Cells, 2001. **19**(4): p. 271-8.
47. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
48. Ferri, A.L., et al., *Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain.* Development, 2004. **131**(15): p. 3805-19.
49. Shields, J.M., R.J. Christy, and V.W. Yang, *Identification and characterization of a gene encoding a gut-enriched Kruppel-like factor expressed during growth arrest.* J Biol Chem, 1996. **271**(33): p. 20009-17.
50. Dang, C.V., et al., *The c-Myc target gene network.* Semin Cancer Biol, 2006. **16**(4): p. 253-64.
51. Gidekel, S., et al., *Oct-3/4 is a dose-dependent oncogenic fate determinant.* Cancer Cell, 2003. **4**(5): p. 361-70.
52. Dong, C., D. Wilhelm, and P. Koopman, *Sox genes and cancer.* Cytogenet Genome Res, 2004. **105**(2-4): p. 442-7.
53. McConnell, B.B., et al., *The diverse functions of Kruppel-like factors 4 and 5 in epithelial biology and pathobiology.* Bioessays, 2007. **29**(6): p. 549-57.
54. Bhattacharya, B., et al., *Gene expression in human embryonic stem cell lines: unique molecular signature.* Blood, 2004. **103**(8): p. 2956-64.
55. Scheper, W. and S. Copray, *The molecular mechanism of induced pluripotency: a two-stage switch.* Stem Cell Rev, 2009. **5**(3): p. 204-23.
56. Rodda, D.J., et al., *Transcriptional regulation of nanog by OCT4 and SOX2.* J Biol Chem, 2005. **280**(26): p. 24731-7.
57. Wilmut, I., *The first direct reprogramming of adult human fibroblasts.* Cell Stem Cell, 2007. **1**(6): p. 593-4.
58. Varas, F., et al., *Fibroblast-derived induced pluripotent stem cells show no common retroviral vector insertions.* Stem Cells, 2009. **27**(2): p. 300-6.
59. Akutsu, T., S. Miyano, and S. Kuhara, *Identification of genetic networks from a small number of gene expression patterns under the Boolean network model.* Pac Symp Biocomput, 1999: p. 17-28.
60. Chen, T., H.L. He, and G.M. Church, *Modeling gene expression with differential equations.* Pac Symp Biocomput, 1999: p. 29-40.
61. Akutsu, T., S. Miyano, and S. Kuhara, *Inferring qualitative relations in genetic networks and metabolic pathways.* Bioinformatics, 2000. **16**(8): p. 727-34.
62. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli.* Nature, 2000. **403**(6767): p. 339-42.

63. Fussenegger, M., et al., *Streptogramin-based gene regulation systems for mammalian cells.* Nat Biotechnol, 2000. **18**(11): p. 1203-8.
64. Corish, P. and C. Tyler-Smith, *Attenuation of green fluorescent protein half-life in mammalian cells.* Protein Eng, 1999. **12**(12): p. 1035-40.
65. Fussenegger, M., et al., *Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells.* Nat Biotechnol, 1998. **16**(5): p. 468-72.
66. Yusa, K., et al., *Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon.* Nat Methods, 2009. **6**(5): p. 363-9.
67. Tian, J., et al., *Accurate multiplex gene synthesis from programmable DNA microchips.* Nature, 2004. **432**(7020): p. 1050-4.
68. Smith, A.G., et al., *Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides.* Nature, 1988. **336**(6200): p. 688-90.
69. Ying, Q.L., et al., *BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3.* Cell, 2003. **115**(3): p. 281-92.
70. Ellis, J., et al., *Alternative induced pluripotent stem cell characterization criteria for in vitro applications.* Cell Stem Cell, 2009. **4**(3): p. 198-9; author reply 202.
71. Mullick, A., et al., *The cumate gene-switch: a system for regulated expression in mammalian cells.* BMC Biotechnol, 2006. **6**: p. 43.