Controlling Transcription and Cell fate during cellular reprogramming and differentiation

Bastiaan Dekker

Summery

Induced pluripotent stem cells provide a promising platform for regenerative medicine. Expression of four transcription factors, Oct4, Sox2, Klf4 and c-Myc (OSKM), can already induce pluripotency from somatic cells. Nevertheless, introducing OSKM factors are bringing concerns regarding inducing tumorigenesis. However, inactivate genome editing tools (ZF, TALE and Crispr-Cas9) have evolved quickly to control gene expression either directly or indirectly without altering the DNA sequence. In this review, we highlight the different uses of those tools for inducing pluripotency and differentiation. In addition, the different aspects of inducing pluripotency and the barriers preventing induced pluripotency are highlighted. We also provide a perspective on future applications of the technologies.

Keywords: Induced pluripotent stem cells (iPSC), Regulation of gene expression, Crispr-Cas9, Zinc Finger, Transcription-Activator-Like Effector (TALE), Effector domains, Epigenetic barriers

Introduction

Pluripotent stem cells have a promising for regenerative medicine, due to the capacity of self-renewal and the possibility of differentiating in every cell type. *Thomson et al.* were the first in 1998 to isolate out of the inner cell mass of a blastocyst human embryonic stem cells (hESC), which had pluripotent properties[1]. Ethical concerns emerged quickly about the use of hESC, due to the destruction of embryos for hESC isolation. In 2006 Takahashi and Yamanaka reported an alternative approach, by generating induced pluripotent stem cells (iPSC) with mouse cells and human cells in 2007[2,3]. Takahashi and Yamanaka revolutionized the idea of cell therapy and personalized medicine, because of the idea that of easily accessible cells could be reprogrammed towards different cell types. Generating iPSCs involves reprogramming of differentiated somatic cells by ectopic overexpressing of the transcription factors *Oct4, Sox2, Klf4* and *c-Myc,* known as OSKM or Yamanaka factors.

Reprogramming somatic cells to a pluripotent state can be achieved by different approaches. The most commonly used factors to introduce pluripotency are the OSKM factors, however *Klf4* and *c-Myc* can be replaced for *Nanog* and *Lin28*[4]. In addition, *c-Myc* is not even necessary for iPSC generation[5]. Although, *c-Myc* increases and accelerates the reprogramming process[6], *c-Myc* also induces tumorigenesis and should be avoided for clinical application[7]. Similar concerns about tumorigenesis exist with the original described method of introducing the OSKM factor through lentiviral induction, because genomic integration can lead to residual OSKM expression or frameshift mutations. Therefore, other strategies have been developed to avoid genomic integration, such as Sendai virus[8], microRNA[9] and mRNA[10]. However, disadvantages of these strategies are that they are labor intense, especially mRNAs, or the high number of passage before all virus residues are gone. Interestingly, recent advances has given a new possibility to induce pluripotency through endogenous gene regulation. Endogenous gene

regulation can be either directly controlling transcription through effector domains or indirectly through epigenetic modifications, epigenetic editing. Regulating the endogenous genes allows cells to express genes through their own transcription machinery and therefore capable of expressing different isoforms and control gene expression more naturally.

Epigenetic editing or direct gene regulation tools are based on genome editing tools, such as zinc-finger proteins (ZFs), transcription activator-like effectors (TALEs) and clustered regularly interspaced short palindromic repeat (CRISPR) Cas9. These genome editing tools are used as DNA binding domains (DBD), however without nuclease activity. The first epigenetic editing tool was engineered in 2002, the ZF system was fused to a histone methyltransferase causing gene specific repression upon induction of H3K9 methylation[11]. Second, The TALE system is another protein binding DNA system using similar fusion constructs as ZF for gene regulation. Recently, the Crispr-Cas9 system was discovered, this editing tool is guide by a sgRNA to a specific genomic site based on Watson-Crick base pairing[12]. All three DBD are fused to effector domains for either activating or repressing gene expression or modifying the chromatin to regulate gene transcription indirectly. These systems are often referred as artificial transcription factors (ATF). However, even without the effector domains the inactivated cas9 (dCas9) can already block the RNA polymerase, leading to gene suppression[13].

As today, various efficient tools have been developed to regulate gene transcription either directly or indirectly. In this review, we discuss the question how gene regulation tools can be used to generate iPSCs and subsequently the differentiation? To this end, we highlight the direct and indirect regulation of the OSKM to promote iPSC generation, before focusing on the pluripotency network and the barriers preventing reprogramming. Next, we highlight the versatility of gene regulation tools focusing on the cell fate and the current possibilities and the future applications.

Reprogramming induced by Effector Domains

Yamanaka showed that re-activation of OSKM factors was necessary to induce the pluripotent state, subsequently many studies have showed the possibility to re-activate OSKM genes with gene regulation tools or epigenetic editing tools. In general, studies used the DBD's fused to activation domains to re-activate the OSKM genes, because somatic cells would normally suppress the OSKM expression. The most commonly used activation effector domain is the herpes simplex viral protein 16 (VP16). As today, VP16's are multiplied to increase the effect of the activation domain. The most often used multiplied VP16 domain is VP64 (4x VP16), however the domain can also be increased to VP160 (10x) or VP192 (12x). Other used activation domains, but in lesser extent, are the single or multiplied p65 activation domain. The p65 subunit is involved in NF-kappaB pathway to regulate gene expression in mammalian cells. So, the effector domains together with the DBD tools gives a variety of options that is available to regulate endogenous gene expression.

The fusion of the VP16 or p65 activation domains to different DBD tools had differential expression effects on the target genes. To this end, Ji et al. looked extensively at the various activation domains fused to the Zinc-finger system[14]. The zinc-finger proteins targeted the OSKM promotors with either fused to a VP16, VP64, P65 or 2xp65. As expected, the activation activity varies between various activation domains, whereas in most cases the multiplied activation domains work slightly better. Both *Sox2* and *c-Myc* had higher transcription levels

with the VP64 domains, whereas the 2xp65 activation domain increased *Oct4* expression as highest, respectively. On the other hand, *Kfl4* expression was activated as highest when bound by p65[14]. Suggested was that the epigenetic state could influence the interactions of the activation domains and thereby the gene activation. Therefore, it would be worth considering which activation domain is used depending also on the cell type, as the epigenetic state differs between cell types. Similar might account for the different targeting systems fused to the activation domains.

Reprogramming requires high expression of the OSKM factors to overcome potential roadblocks, therefore solely using gene control systems requires sufficient expression to induce pluripotency. One of the main focused genes is the *Oct4* gene, due to its importance in inducing pluripotency (see below). To this matter, Gao et al. used TALE-VP64 targeting the enhancer to activate *Oct4* expression. Despite the fact, that a single TALE-VP64 increased expression only with 4-fold. Multiple TALE-VP64 increased the *Oct4* expression up to 20-fold, although this was in the presence of ectopic co-expressed Sox2, Klf4 and c-Myc. Interestingly, this 20-fold increased *Oct4* expression was sufficient to generate iPS cells[15]. Similarly, Hu et al. increased endogenous *Oct4* expression up to 20-fold with multiple TALE-VP64 as well, however instead of the enhancer the promotor was targeted. In addition, dCAs9-VP64 targeting also the promotor increased the endogenous *Oct4* expression up to 17-fold[16]. Notable, both TALE and dCas9 achieved similar Oct4 expression as Gao et al. reported, however both without the use of SKM factors. Leaving up for debate if the SKM factors had a synergetic effect on *Oct4* expression or if the promotor can be better targeted to achieve higher expression rates. Furthermore, Gao et al. did generate iPS cells, whereas Hu et al. did not by omitting the other Yamanaka factors[15,16]. Interestingly, Hu et al. observed that *Oct4* expression can be increased but not sustained, suggesting that TALE or dCAs9 could not overcome the epigenetic barrier[16]. So, although *Oct4* is important in reprogramming, *Oct4* activation alone was not sufficient to overcome the epigenetic repression. In fact, Balboa et al. showed that *Oct4* expression could be upregulated to 70-fold with dCas9VP192 targeted with five gRNAs. However, 70-fold increased *Oct4* expression was still not sufficient, because pluripotency was only induced upon co-expression of Sox2 and KLf4[17]. Nevertheless, Balboa et al. used controllable dCas9VP192 upon treatment with Trimethoprim, gaining the advantage of controlling gene expression even more which might give more insight in the reprogramming events and differentiation. These studies showed promising evidence that endogenous *Oct4* expression can be increased sufficiently for reprogramming, however other factors still need to be co-expressed for inducing the pluripotency.

Activating *Oct4* expression can replace ectopic expression for reprogramming, yet solely *Oct4* expression is not sufficient therefore have *Sox2*, *Klf4* and *c-Myc* been reviewed as well. In addition, to Ji et al. that used the ZF system, the TALE and Crispr-Cas9 systems have been used to upregulate *Sox2* and *Klf4* as well. Zhang et al. used the TALE-VP64 system to activate endogenous *Sox2* and *Klf4* expression. The Sox2-TALE and Klf4-TALE were able to upregulate expression with 5.5-fold for Sox2 and 2.2-fold for Klf4[18]. However, both genes were only targeted by one TALE-VP64 at the same time[18]. Therefore, gene expression changes might be lower. Cheng et al. used dCas9-VP160 to activate *Sox2* expression by targeting with three sgRNAs the promotor. The three sgRNAs together caused a synergistic effect upregulating the Sox2 expression with 7-fold, whereas one single sgRNA did not had any effect[19]. Interestingly,

Cheng et al. targeted the *Oct4* promotor as well with a similar system, however only an 8-fold increased expression was measured[19]. In contrast, Hu et al. targeted the same promotor with four sgRNA dCAs9-VP64 resulting in a 20-fold increased *Oct4* expression[16]. Possible explanation is the different activation domains fused to the dCas9, as previous discussed (see above). Another explanation is that dCas9 downstream the promotor could lead to steric hindrance of the transcription. Cheng et al. observed that dCas9-VP160 targeted downstream of the transcription start site (TSS) negatively affected the synergistic effect to activate gene expression[19]. This suggesting that not only the activation domain but also the targeting site of the DBDs affect the expression rates.

Although, most studies have studied activation of the OSKM genes to induce pluripotency, repression of genes might give information about inducing pluripotency as well. Zhang et al. fused the Kruppel-associated box (KRAB) to TALE in order to suppress the miR-302/367 cluster. The TALE-KRAB system targeted the promotor causing a 70 – 90% knockdown of the miR-302/367 cluster. Consequently, iPSC generation was inhibited upon OSKM factor induction[20]. In addition, overexpressing the miR-302/367 cluster enhanced reprogramming, as *Nanog* repressor MBD2 expression was suppressed[21]. Nevertheless, the miR-302/367 cluster only enhances iPSC generation, no evidence is found that it can be substituted for any of the OSKM factors. Surprisingly, KRAB is traditionally reported as transcriptional repressor, however ZF-KRAB also activated the *Oct4* gene in ovarian and breast cancer cell lines[22]. Despite the fact, that both cancer cell lines are associated with upregulated *Oct4* expression, this might also be an effect of the engineered ZF-KRAB protein[23,24].

Gene activation is not only limited to fused activation domain, in fact epigenetic modifying enzymes fused to DNA binding domains can be just as effective. These tools are referred as epigenetic editing tools, as they change the epigenome. Hu et al. reported that activation of *Oct4* alone through activation domains might not be able overcome the epigenetic barrier[16]. However, there is also evidence showing that VP16 can recruit histon acetyltransferase (HAT) to acetylate histon marks, which are strongly associated with gene activation[25]. Nevertheless, activation domains remodel the chromatin indirectly and may therefore be less potent. To this end, Hilton et al. engineered a dCAs9-P300, a conserved acetyltransferase, leading to H2K27ac. Their data showed that the dCas9-P300 upregulates *Oct4* expression with a 32-fold, whereas similar dCas9-VP64 achieved 20-fold upregulation. More interesting, activation through p300 was already be achieved by one gRNA instead of pooling five or six gRNAs[26]. Based on these observations it could be suggested that other OSKM genes could be targeted as well, as both *Sox2* and *Klf4* only achieved a modest 7-fold upregulation. Moreover, combination of dCAs9-VP64 and dCas9-P300 with multiple sgRNAs could be even more beneficial, because transcription will be already induced and the epigenetic barrier might be by-passed for further reprogramming.

These studies show promising results regarding artificial transcription factors regulating gene expression to promote iPSC generation. However, so far no study reported be capable of generating iPSC solely using the ATFs. Nevertheless, various studies have shown that the OSKM genes needed to induce pluripotency can be regulated. However, considered should be the level of increased expression. As both *Sox2* and *Klf4* expression was only 7-fold increased, whereas *Oct4* can be increased 70-fold. Interestingly, Wang et al. fused the VP16 domain to Oct4, Sox2 and Nanog, respectively. These engineered factors were not only able to generate iPS cells, but

also enhance the efficiency. Strikingly, Oct4-VP16 alone was already sufficient to introduce pluripotency[27]. So, despite the contrary reports if VP16 alone can recruit histon modifying enzymes, in combination with Oct4 it seems to work fine[27]. This evidence supports even more the suggestion to the target OSKM genes with both VP16 and P300 for increased efficiencies and overcoming possible epigenetic barriers. Especially, for *Sox2* and *Klf4* due to a lower fold increased expression. However, expression rates should not be exaggerated, as overexpression can lead to differentiation as well.

Importance of the Yamanaka Factors in Pluripotency

Reprogramming somatic cells requires only four transcription factors, OSKM, resulting in changes in the transcriptional and chromatin state. However, mechanisms underlying the molecular events of reprogramming are still unclear. Many questions remain about: how the OSKM factors can introduce reprogramming mechanisms and alter the epigenetic state ?

Pluripotency Network: Oct4 and Sox2

Although alternative sets of transcription factors have been identified to promote iPSC generation, most of them include *Oct4* and *Sox2*. The common interest in expressing *Oct4* and *Sox2* might be due to the fact that these proteins are identified, together with a third gene *Nanog*, as key factors for maintaining the pluripotency. The pivotal role of *Oct4, Sox2* and *Nanog* has been reviewed in several studies. For example, *Oct4*-deficient embryos were impaired to develop and restricted to differentiate into trophectoderm[28]. In fact, embryonic stem cells with decreased or increased *Oct4* expression resulted in differentiation and loss of pluripotency[29,30]. In 2003, Nanog was discovered as a transcription factor maintaining the pluripotency, as *Nanog* deficient stem cells lost their pluripotency[31]. Reduced *Sox2* expression in embryonic stem cell resulted in differentiation towards trophectoderm. In addition, loss in pluripotency was shown as well, similar to the *Oct4* deficient embryonic stem cells. Moreover, knocking down *Sox2* affected also the *Oct4* and *Nanog* expression, as both were reduced as well. Suggesting that *Oct4, Sox2* and *Nanog* are linked in maintaining the pluripotent state[32].

Similar results in knockout cells between Oct4 and *Sox2* was no coincidence, because both transcription factors co-bind target genes as heterodimer. Moreover *Oct4*, *Sox2* and *Nanog* have all shared target genes including themselves. Those target genes have been identified with chromatin immunoprecipitation (Chip) sequencing, showing that *Oct4*, *Sox2* and *Nanog* co-occupy various active, but also silenced target sites[33,34]. In addition, binding together to the promotors of their own genes, autoregulatory loops can be formed to maintain the pluripotency (see Fig 1). Binding activated or silenced target sites, *Oct4*, *Sox2* and *Nanog* can promote expression of pluripotency associated factors and repress lineage specific genes.

Pluripotency Network: Klf4 and c-Myc

Maintaining pluripotency is mainly regulated by *Oct4* and *Sox2*, whereas *Klf4* and *c-Myc* are important to acquire the pluripotent state. Although, the role of *Klf4* and *c-Myc* are rather associated with the induction then maintaining the pluripotent state, both transcription factors do facilitate the pluripotent state. For instance, *Klf4* targets the promotor of *Nanog* and therefore enhances the effect of *Oct4* and *Sox2* to maintain *Nanog* expression and therefore pluripotency[35]. However, *Klf4* knockout mice developed normally without phenotypically and

histologically abnormalities, nevertheless skin barrier dysfunction caused a lethality shortly after birth[36].

c-Myc knockout embryonic stem cells developed normally at early stages, but 10.5 days after gestation mouse models died due to abnormalities in heart and vascular tissue[37]. So, as *c-Myc* is not necessary for iPSC generation and the embryo's survived at early stages, hence it was suggested that *c-Myc* is not crucial for pluripotency[5,37]. However, Smith et al. showed that either *c-Myc* or *n-Myc* is necessary for maintaining pluripotency and that loss of both *c-Myc* and *n-Myc* causes iPSCs to differentiate towards endodermal lineages. Normally c-Myc would represses GATA6, master regulator that drives endoderm differentiation[38]. Taken together, both *Klf4* and *c-Myc* might be not be crucial or indispensable for iPSC generation or maintaining pluripotency, yet both do facilitate the maintenance of the pluripotency network.

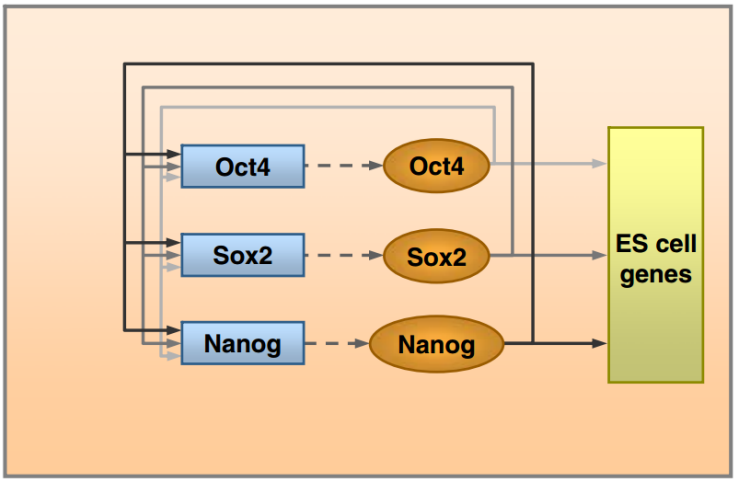

Figure 1: Oct4, Sox2 and Nanog form a autoregulatory circuit. All three genes target their own as well as the others promotor. Thereby stably maintaining their own and the others expressions, to contribute to maintaining the pluripotency[70].

The Role of Yamanaka Factors in Reprogramming

The acquisition of pluripotency by reprogramming through expression of OSKM depends not only on transcriptional changes but is also accompanied by epigenetic remodeling. Reports showed that reprogramming of somatic cells requires two transcriptional waves. Polo et al. defined that the first transcriptional wave is mediated by *c-Myc* and *Klf4*, whereas the second wave is driven by *Oct4*, *Sox2* and *Klf4*[39]. Initially, in the first phase genes involved in proliferation and metabolism are upregulated, whereas genes involved in development are downregulated (see Fig. 2). Subsequently, in the second phase genes associated with pluripotency and embryonic development are upregulated. Between the first and second wave cells remain in an intermediate state where they can become refractory for inducing pluripotency, which might explain the low efficiency[39]. Although, molecular mechanism are yet unclear, developmentally silenced genes involved in pluripotency have to be reactivated. These silenced genes are mainly found in heterochromatin, suggesting that transcription factors used for reprogramming can alter the chromatin state to eventually express those genes. Transcription factors with the ability of accessing tightly packed chromatin and initiating chromatin remodeling are referred as "Pioneer Factors". Interestingly, reports showed that only *Oct4, Sox2* and *Klf4* are pioneer factors[40].

The first wave

The first wave was initiated through *c-Myc* and *Klf4* to upregulate proliferation genes and repress developmental genes. Although, both *c-Myc* and *Klf4* are dispensable, still *Klf4* was shown to be a pioneer factor. The role of *c-Myc* in reprogramming was initially thought to enhance and accelerate the efficiency, through promoting cell proliferation[6]. In addition, in the absence of ectopic *c-Myc* the efficiency was drastically reduced[5], but besides reprogramming was delayed as well[6]. Therefore, proliferation might be a contributing factor in reprogramming events, as cellular division provides an opportunity to remodel chromatin and re-establish transcriptional programs. Another possibility is that c-Myc recruits chromatin remodelers, including SWI/SNF nucleosome remodelers[41], histon acetyltransferase P300[42] and histon demethylases/mytheltransferases[43,44]. However, a key feature of pioneer factors is the recruitment of chromatin modifiers. Therefore, OSK do not need c-Myc for reprogramming, moreover c-Myc does not bind closed chromatin on its own[45]. In fact, c-Myc is only capable of binding closed chromatin in the presences of the OSK factors. Suggesting that c-myc co-binds with OSK to closed chromatin and thereby facilitates reprogramming and enhances the efficiency. Supporting this is the finding that VP16 activation domains fused to Oct4, Sox2 and Nanog increases the reprogramming efficiency[27]. As VP16 has the ability to recruit other transcription factors and chromatin modifiers. So, during reprogramming the OSK factors might not function optimally to overcome epigenetic barriers and c-Myc facilitates this process. More recent data suggest that c-Myc does not target unique targets within the pluripotency network, but rather acts as an amplifier of transcription possibly through the mechanism of Pol II pause releasing[46,47]. The role for *Klf4* during the first wave was predominately repressing somatic target genes. Polo et al. showed that Klf4 downregulates many of his target genes during the first wave, whereas c-myc upregulates his target genes[39].

The second wave

The second wave is mediated by *Oct4, Sox2* and *Klf4* to upregulate pluripotency genes, including *Nanog*, and eventually establish the pluripotency network. All three factors are proposed to be pioneer factors. Although the mechanism that involves chromatin remodeling through pioneer factors is not fully understood, it is shown that Oct4, Sox2 and Klf4 bind closed chromatin upon expression in somatic cells[40]. Suggested was that pioneer factors can recruit chromatin modifiers. In fact, Oct4 can activate histon demethylase genes, *Jmjd1a* and *Jmjd2c*[48]. Both Jmjd1a *and* Jmjd2a act as H3K9 demethylases, H3K9me3 mark is associated with repression of genes, including *Nanog*. Supporting to this evidence is that reprogrammed *Jmjd1a* and *Jmjd2c* knock down cells remained in the intermediate state that Polo et al. described earlier[39,49]. Another Oct4 and Sox2 associated chromatin modifier is the SWI/SNF complex. Interestingly, overexpression proteins involved in the SWI/SNF complex enhances reprogramming mediated by Oct4, Sox2 and Klf4[50,51].

The dual role of Klf4 proposed by Polo et al. indicates that Klf4 supports gene repression during the first phase and enhances the pluripotency gene expression at the second phase[39]. For example, *Nanog* is not necessary for reprogramming, whereas it is for maintaining the pluripotency network. Besides Oct4 and Sox2 targets Klf4 the promotor of *Nanog* as well[35]. Furthermore, Klf4 interacts with CDK9, a RNA pause releasing proteins, which is in contrast with the previous suggested idea that c-Myc promotes pause releasing[52]. However, c-Myc targets are

more associated with proliferation, suggesting that c-Myc can act as pause releasing factor but only during phase one. Supporting this observation is the finding that inhibition of Klf4 at later phases of reprogramming, cells remained in the intermediate state[53]. Moreover, similar results were reported when CDK9 was inhibited[52]. Suggesting that both CDK9 and Klf4 are important for reprogramming at the late phase. Besides effecting directly the gene regulation, Klf4 was also necessary for long-range promotor-enhancer looping. Wei et al. Reported that deletion of Klf4 causes disruption of the long-range interaction and a decrease Oct4 expression[54]. Taken together, these data indicates that Klf4 acts as an activator during the second wave.

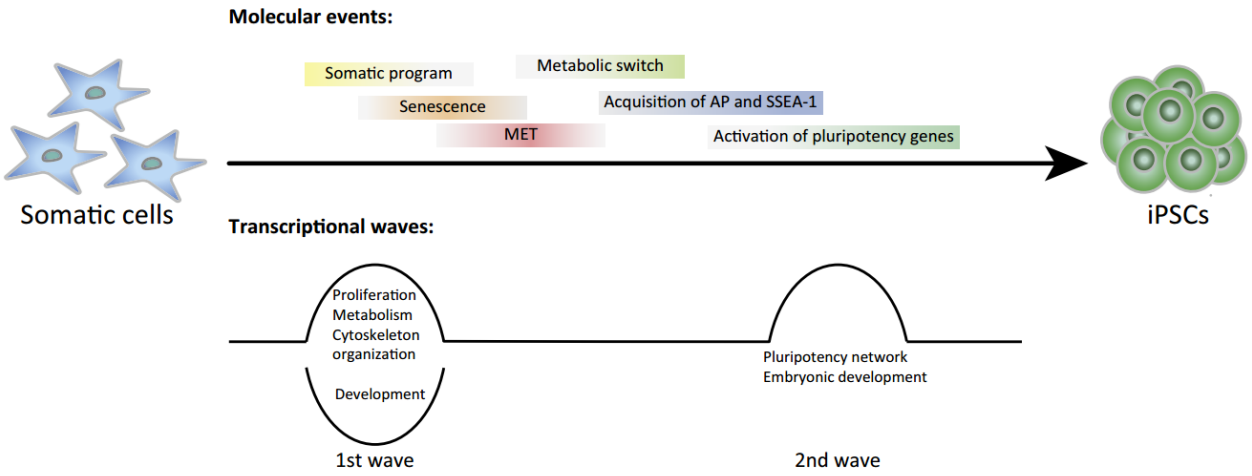

Figure 2: Schematic overview of reprogramming waves inducing pluripotency. First wave is mediated by *Klf4* and *c-Myc* to upregulated the metabolism and proliferation, wereas developmental genes are downregulated. Second wave is mediated by *Oct4, Sox2* and *Klf4* to induce the pluripotency network and maintain this. Between the transcriptional waves cells might become refractory for inducing pluripotency and subsequently differentiate[71].

Reprogramming Barriers

Although reprogramming of somatic cells towards iPS cells is a widely used technique, it still is a time-consuming and low efficiency process, indicating possible resistance towards reprogramming. Overcoming these shortcomings many studies are focused on altering the epigenetic regulation. For example, methyltransferases causing gene repression have been suggested to inhibit reprogramming efficiencies. Speculated was that inhibition of histon methyltransferase expression would enhance the efficiency. Indeed, histon methyltranserase DOT1L knock down cell lines accelerated the reprogramming and increased the yield of iPSC colonies. Moreover, suppressed *DOT1L* expression could substitute ectopic *Klf4* and *c-Myc*, as both *Nanog* and *Lin-28* were upregulated early phase[55]. Besides the benefits of accelerated and increased reprogramming yields, the exclusion of *c-Myc* makes it even more interesting due to the tumorigenesis effect of c-Myc. Similarly, increased expression of histon demethylase enzymes enhances the reprogramming. Overexpressing Jmjd1a *and* Jmjd2a, recruited by Oct4 (see above), increased the iPSC generation[49]. However, overexpressing or repression of addition genes, besides OSKM, might bring some concerns with it. Therefore, chemical approaches might be preferred. In addition to the histon demethylases, vitamin C is a chemical compound that promotes histone demethylase activities of Jhdm1a/1b and thereby accelerating and enhancing reprogramming[56,57].

Furthermore, changes in the acetylation and deacetylation may affect the reprogramming efficiency as well. Generally, deacetylation is associated with gene repression, whereas acetylation is associated with gene activation. Deacetylation and nucleosome remodeling complex, NuRD, is recruited by OSKM factors, through interaction with the core component Mbd3. Depletion of the Mbd3 protein resulted in a striking 95% reprogramming efficiency, whereas normal cells had 18% efficiency[58]. Normally, *Mbd3* is only expressed in somatic cells, but not in embryonic stem cells. Therefore, *Mbd3* expression is a roadblock for reprogramming of somatic cells.

Another difference between somatic cell and embryonic stem cell expression is the macroH2A. MacroH2A deposition is normally enriched in somatic cells, as these histon variant is associated with heterochromatin, however embryonic stem cells are considered have a more "open" chromatin. MacroH2A has two distinct forms encoded in two genes, *H2afy* and *H2afy2*. Knocking down of both genes showed that reprogramming was enhanced, suggesting that the MarcoH2A acts as an epigenetic barrier for reprogramming[59,60]. However, many studies have focused increasing the efficiency but in most cases this is related with overexpressing or suppressing additional genes. Moreover, the effect of overexpression or repression has to be temporary, as for example the Mdb3 or MacroH2A proteins are important for somatic cells. Inducible expression system for either gene expression or RNAi might be suitable. However, chemical compounds may be preferred, as it is easy to administer or remove from the cells, without the need to transfect vectors, mRNAs or miRNAs. Nevertheless, these molecular mechanisms might improve the safety anyway, due to the fact that conventional transcription factors use *c-Myc* to generate iPSCs.

Cell Differentiation induced by Effector Domains

Regulating gene transcription through effector domains have a broad spectrum of applications, besides inducing pluripotency also differentiation of iPS cells or transdifferentiation can be achieved. Targeting the *Oct4*, *Sox2* and *Klf4* genes to induce pluripotency can also affect differentiation. Reprogramming is inefficient and takes long, therefore reprogrammed intermediates cells (pre-IPSC) might dedifferentiate towards other cell lineages[39]. This possibly due to aberrant expressed OSKM genes or their target genes. For example, 2-fold upregulation of *Oct4* causes differentiation towards endoderm or mesoderm, whereas overexpression of *Sox2* causes differentiation towards neuroectoderm or mesoderm[61,62]. These observations indicate that *Oct4* and *Sox2* expression need to be maintained between limits. However, overexpression can also be beneficial when iPSCs need to be differentiated. Downside is that both Oct4 and Sox2 do not differentiate towards specific cell types, therefore cell type specific transcription factors are needed.

Transcription factors leading differentiation toward lineage specific cell types do not require much additional activated genes, moreover one transcription factor can already be sufficient. Development of neural tissue differentiated from iPSC cells can be promoted by either neurogenin 2 (NGN2) or neurogenic differentiation factor 1 (NeuroD1). Targeting either NGN2 or NeuroD1 with multiple gRNA dCas9-VPR (combination of VP16, P65 and Rta activation domains) showed that after 4 days a neuronal phenotype was observed[63]. However, targeting more genes at once might be beneficial. Balboa et al. promoted differentiation towards pancreatic cells, by simultaneously targeting endodermal transcription factors *Foxa2* and *Sox17*

and pancreatic transcription factors *Pdx1* and *Nkx6.1*. Targeting both endodermal and pancreatic transcription factors increased the differentiation time from 10 days up to 3 days[17]. However, not only activation can cause differentiation, repression can lead also to differentiation as well. Yet, repression is only interesting if genes are already expressed, therefore pluripotent genes are the most interesting ones. Kearns et al. showed that downregulating *Oct4* or *Nanog* causes stem cells to differentiate, but not to specific cell types[64]. Therefore, gene repression is rather supportive to cell type specific differentiation then to initiating cell type specific differentiation.

Instead of directly regulating gene expression to induce differentiation, chromatin modifying enzymes have been used as well. For example, LSD1 a demethylase that is associated with gene repression. LSD1 has been fused with TALE and dCas9 to target *Oct4*. Results showed reduction of H3K4me2 and H3K27ac, which leads to reduced *Oct4* expression. However, similar as for the repression domains, these chromatin modifications can only support cell type specific differentiation. Chromatin remodelers causing gene activation on the other hand are capable of inducing differentiation. In fact, Hilton et al showed that dCas9-p300 can increase *MyoD1* expression up to 50-fold, whereas dCas9-VP64 had only 25-fold upregulation. Although, Hilton et al. did not show that this upregulation can lead to differentiation others did[26]. MyoD1, myogenic differentiation 1, is a transcription factor pivotal for differentiation towards myogenic tissue[65]. Interestingly, upregulation of MyoD1 with dCAs9-VP64 is already sufficient to achieve transdifferentiation. Chakraborty et al. showed that fibroblasts with endogenous activated *MyoD1* lead to direct conversion toward skeletal myocytes[66]. This all shows the versatility of regulating gene repression but certainly gene activation through effector domains fused to DBDs

Future perspectives

In recent years, systems for control gene expression have been developed through the fusion of effector domains with DNA binding domains. The three systems, ZF, Talen and Crispr-Cas9 have been used to either directly control gene regulation on transcriptional level, or fused with histon modifier enzymes to regulate gene transcription indirectly on epigenetic level. These fusion constructs have shown promising results in the approaches of generating iPSCs and the differentiation.

The strengths of the endogenous gene regulation are pointed out by the examples listed above, showing that endogenous gene regulation is a versatile tool with possible applications in the near future. Besides the generation and differentiation of iPS cells, endogenous gene regulation may be applied for cancer treatment or preventing fibrosis in transplanted tissue as well. Looking in particular to the iPSC generation and differentiation, the regulation of gene expression might be more suitable then other methods. Integration-free iPSC generation with Oct4-VP16 has shown to be suitable[27], however the induced expression is uncontrollable. Moreover, systems requiring transferred cDNA are limited to the size and isoform as well. In contrast, with activation domain and especially with epigenetic editing a more natural gene expression could be induced, consisting out of various isoforms. More interestingly, the epigenetic editing tools could be suitable for “hit and run” approaches, by targeting the chromatin to generate sustained gene silencing or activation.

The biggest roadblock for reprogramming seems to be the epigenetic barrier, leading to long culture periods and low efficiencies. Chemical compounds increasing the efficiency are interesting, due to the fact that these compounds can be easily administered or removed from the cells, consistent with the "hit and run" approach. However, "hit and run" approaches, like epigenetic editing, should induce sufficient expression otherwise cells might remain in the intermediate state and become refractory, leading to possible transdifferentiation[39]. A possible advantage of the "hit and run" approach is to target *c-Myc*. As enhanced iPSC generation is shown in the presence on *c-Myc* activation, but ectopic expression may lead to tumorigenesis[7]. However, with "hit and run" approaches *c-Myc* can be target during reprogramming and during differentiation the gene can be naturally repressed again.

Obviously, all of the DNA binding system has their own limitation, however the recent advances with the inactivated Cas9 has created a powerful tool for regulating gene expression. The zinc-finger proteins might have as advantage to be the least immunogenicity due to the foreign mammalian origin of both TALE and Crispr. However, potential limitation of the three targeting systems is the accessibility to the promotor or enhancer, as specific DNA motifs are required for DNA binding. So, the accessibility depends on the DNA sequence, whereas targeting to much downstream of the TSS could lead to steric hindrance.

Recent advances, has made particular interest in the dCas9 method due to the ease of targeting various targets, moreover the dCas9 provides an easy method to target with multiple sgRNAs the same gene. Especially, as multiple targets are often needed to generate sufficient expression. In contrast, ZFs and TALEs are more difficult to design and engineer, especially for multiple targets. Recently, one of the potential limitation of off-target effects have been reduced drastically. The "enhanced specificity" spCas9 has mutation reducing possible mismatching[67]. Another advancement raised much interest as well. The Cas9 can be introduced in the cell without plasmid or virus, by directly bringing the protein inside the cell through deformation techniques. This gives the possibility to introduce Cas9 even more safer with a high efficiency, moreover cas9 can also be introduced in embryonic stem cells [68]. Another recently published technique made effort of endocytosis to introduce dCas9, in addition this system can be applied for TALE as well[69]. These could bring the "hit and run" approach even closer. In addition, activation domains might be used then as well, due to the fact that Cas9 degrades within cells. However, this might become a time-consuming technique. Therefore, future studies should investigate these intriguing possibilities to promote iPSC solely using endogenous gene regulation tools. Taken together, all of these observations supports the idea of how versatile these tools are and the possibility for future application.

Reference

1. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (80-.).* **282,** 1145–1147 (1998).
2. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).
3. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131,** 861–872 (2007).
4. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science (80-.).* **318,** 1917–1920 (2007).
5. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. Correspondence c-Myc Is Dispensable for Direct Reprogramming of Mouse Fibroblasts. *Cell Stem Cell* **2,** 10–12 (2008).
6. Sridharan, R., Tchieu, J., Mason, M. J., Yachechko, R. & Kuoy, E. Resource Role of the Murine Reprogramming Factors in the Induction of Pluripotency. *Cell* **136,** 364–377 (2009).
7. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–318 (2007).
8. Ban, H. *et al.* Ef fi cient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *PNAS* **108,** 2–7 (2011).
9. Anokye-danso, F. *et al.* Article Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Stem Cell* **8,** 376–388 (2011).
10. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).
11. Snowden, A. W. *et al.* Gene-Specific Targeting of H3K9 Methylation Is Sufficient for Initiating Repression In Vivo. *Curr. Biol.* **12,** 2159–2166 (2002).
12. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
13. Gilbert, L. A. *et al.* Resource CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2012).
14. Ji, Q. *et al.* Engineered zinc-finger transcription factors activate OCT4 (POU5F1), SOX2 , KLF4 , c-MYC (MYC) and miR302 / 367. *Nucleic acids Reseaerch* **42,** 6158–6167 (2014).
15. Gao, X. *et al.* Reprogramming to Pluripotency Using Designer TALE Transcirption Factors Targeting Enhancers. *Stem Cell Reports* **1,** 183–197 (2013).
16. Hu, J. *et al.* Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic acids Reseaerch* **42,** 4375–4390 (2014).
17. Balboa, D. *et al.* Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. *Stem Cell Reports* **5,** 448–459 (2015).
18. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29,** 149–154 (2011).
19. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–71 (2013).
20. Zhang, Z. *et al.* Dissecting the Roles of miR-302/367 Cluster in Cellular Reprogramming Using Tale-based Repressor and Talen. *Stem Cell Reports* **1,** 218–225 (2013).
21. Lee, M. R. *et al.* Epigenetic Regulation of NAnog by MiR-302 Cluster-MBD2 Completes Induced Pluripotent STem Cell Reprogramming. *Stem Cell* **31,** 666–681 (2013).
22. Juárez-moreno, K. *et al.* Breaking through an epigenetic wall Re-activation of Oct4 by KRAB-containing designer zinc finger transcription factors. *Epigenetics* **8,** 164–176 (2013).
23. Beltran, A. S. *et al.* Generation of tumor-initiating cells by exogenous delivery of OCT4 transcription factor. *Breast Cancer Res.* **13,** R94 (2011).
24. Zhang, Z. *et al.* Follicle-stimulating hormone inhibits apoptosis in ovarian cancer cells by

regulating the OCT4 stem cell signaling pathway. *Int. J. Oncol.* **43,** 1194–1204 (2013).
25. Memedula, S. & Belmont, A. S. Sequential recruitment of HAT and SWI/SNF components to condensed chromatin by VP16. *Curr. Biol.* **13,** 241–246 (2003).
26. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** 510–7 (2015).
27. Wang, Y. *et al.* Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep* **12,** 373–378 (2011).
28. Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* **95,** 379–391 (1998).
29. Gidekel, S., Pizov, G., Bergman, Y. & Pikarsky, E. Oct-3 / 4 is a dose-dependent oncogenic fate determinant. *Cancer Cell* **4,** 361–370 (2003).
30. Zaehres, H. *et al.* High-efficiency RNA interference in human embryonic stem cells. *Stem Cell* **23,** 299–305 (2005).
31. Mitsui, K. *et al.* The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113,** 631–642 (2003).
32. Fong, H., Hohenstein, K. a & Donovan, P. J. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem Cells* **26,** 1931–8 (2008).
33. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
34. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38,** 431–440 (2006).
35. Zhang, P., Andrianakos, R., Yang, Y., Liu, C. & Lu, W. Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285,** 9180–9189 (2010).
36. Segre, J. a, Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat. Genet.* **22,** 356–360 (1999).
37. Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. & Bradley, A. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.* **7,** 671–682 (1993).
38. Smith, K. N., Singh, A. M. & Dalton, S. Myc represses primitive endoderm differentiation in pluripotent stem cells. *Cell Stem Cell* **7,** 343–354 (2010).
39. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
40. Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161,** 555–568 (2015).
41. Cheng, S. W. *et al.* c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat. Genet.* **22,** 102–105 (1999).
42. Vervoorts, J. *et al.* Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. *EMBO Rep.* **4,** 484–490 (2003).
43. Amente, S. *et al.* LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene* **29,** 3691–702 (2010).
44. Brenner, C. *et al.* Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **247600509,** 336–346 (2005).
45. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).
46. Nie, Z. *et al.* c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* **151,** 68–79 (2012).
47. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).

48. Loh, Y., Zhang, W., Chen, X., George, J. & Ng, H. Demethylases Regulate Self-Renewal in Embryonic Stem Cells. 2545–2557 (2007). doi:10.1101/gad.1588207.namics
49. Chen, J. J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45,** 34–42 (2013).
50. Lessard, J. A. & Crabtree, G. R. Chromatin regulatory mechanisms in pluripotency. *Annu. Rev. Cell Dev. Biol.* **26,** 503–32 (2010).
51. Singhal, N. *et al.* Chromatin-remodeling components of the baf complex facilitate reprogramming. *Cell* **141,** 943–955 (2010).
52. Liu, L. *et al.* Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15,** 574–588 (2014).
53. Nishimura, K. *et al.* Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. *Stem Cell Reports* **3,** 915–929 (2014).
54. Wei, Z. *et al.* Klf4 organizes long-range chromosomal interactions with the OCT4 locus inreprogramming andpluripotency. *Cell Stem Cell* **13,** 36–47 (2013).
55. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
56. Esteban, M. A. *et al.* Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell* **6,** 71–79 (2010).
57. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9,** 575–587 (2011).
58. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).
59. Gaspar-Maia, A. *et al.* MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nat. Commun.* **4,** 1565 (2013).
60. Pasque, V. *et al.* Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. *J. Cell Sci.* **2,** 6094–6104 (2012).
61. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24,** 372–376 (2000).
62. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26,** 903–911 (2008).
63. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12,** 326–8 (2015).
64. Kearns, N. a *et al.* Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141,** 219–23 (2014).
65. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132,** 2685–2695 (2005).
66. Chakraborty, S. *et al.* A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Reports* **3,** 940–947 (2014).
67. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351,** 84–88 (2015).
68. Han, X. *et al.* CRISPR-Cas9 delivery to hard-to-transfect cells via membrane deformation. *Sci. Adv.* **1,** e1500454 (2015).
69. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33,** 73–80 (2014).
70. Scheper, W. & Copray, S. The molecular mechanism of induced pluripotency: A two-stage switch. *Stem Cell Rev. Reports* **5,** 204–223 (2009).
71. Xu, Y. *et al.* Transcriptional Control of Somatic Cell Reprogramming. *Trends Cell Biol.* **xx,** 1–17 (2016).