



Open Access

INVITED REVIEW

Germ Cell Biology

# Reprogramming cells with synthetic proteins

Xiaoxiao Yang\*, Vikas Malik\*, Ralf Jauch

Conversion of one cell type into another cell type by forcibly expressing specific cocktails of transcription factors (TFs) has demonstrated that cell fates are not fixed and that cellular differentiation can be a two-way street with many intersections. These experiments also illustrated the sweeping potential of TFs to “read” genetically hardwired regulatory information even in cells where they are not normally expressed and to access and open up tightly packed chromatin to execute gene expression programs. Cellular reprogramming enables the modeling of diseases in a dish, to test the efficacy and toxicity of drugs in patient-derived cells and ultimately, could enable cell-based therapies to cure degenerative diseases. Yet, producing terminally differentiated cells that fully resemble their *in vivo* counterparts in sufficient quantities is still an unmet clinical need. While efforts are being made to reprogram cells nongenetically by using drug-like molecules, defined TF cocktails still dominate reprogramming protocols. Therefore, the optimization of TFs by protein engineering has emerged as a strategy to enhance reprogramming to produce functional, stable and safe cells for regenerative biomedicine. Engineering approaches focused on Oct4, MyoD, Sox17, Nanog and Mef2c and range from chimeric TFs with added transactivation domains, designer transcription activator-like effectors to activate endogenous TFs to reprogramming TFs with rationally engineered DNA recognition principles. Possibly, applying the complete toolkit of protein design to cellular reprogramming can help to remove the hurdles that, thus far, impeded the clinical use of cells derived from reprogramming technologies.

*Asian Journal of Andrology* (2015) 17, 394–402; doi: 10.4103/1008-682X.145433; published online: 09 January 2015

**Keywords:** cellular reprogramming; protein design; protein engineering; synthetic transcription factors; transactivation

## SWITCHING CELL FATES WITH TRANSCRIPTION FACTOR PROTEINS

The notion that organismic development does not stubbornly follow a predetermined path but is a rather plastic process has been made long before DNA was recognized as the carrier of inheritable information.<sup>1</sup> Nearly 100 years later, cellular reprogramming was first demonstrated when live tadpoles arose after transferring the nuclei of frog intestine derived somatic cells into oocytes.<sup>2</sup> This observation led to the realization that even fully differentiated cells must still contain the complete genetic blueprint required to build a whole organism. That transcription factors (TFs) are remarkably powerful to drive cellular reprogramming, that is to convert one cell type into another, was first demonstrated by turning a mouse fibroblast into a muscle cell with a cDNA encoding a single TF, MyoD (Myod1).<sup>3</sup> Subsequently, a cocktail of four TFs, Krüppel-like factor 4 (Klf4), c-Myc, Sry (sex determining region y) box2 (Sox2) and octamer binding protein 4 (Oct4), was discovered to induce pluripotent stem cells (iPSCs) when forcibly expressed in fibroblasts of mouse<sup>4</sup> and human.<sup>5,6</sup> Since this seminal work, cellular reprogramming has become a mainstream research activity. Like embryonic stem cells derived from the blastocyst, iPSCs can be passaged in culture for indefinite periods and given appropriate growth conditions, can be differentiated into all cell types of the body.<sup>7</sup> The latter can straightforwardly be demonstrated by transplanting iPSCs back into blastocysts, but it is a challenge to recapitulate this process *in vitro*.

The MyoD experiment already demonstrated that differentiated cells do not have to be pushed up all the way the “Waddington canal”<sup>8</sup> to a completely undifferentiated cell type before subsequent re-specification to an alternative cell fate. Instead, the MyoD conversion of fibroblasts directly to muscle indicates that direct transdifferentiation can be accomplished without a pluripotent intermediate. Yet, analogous to the MyoD example, lineage conversion was initially only reported between cell types originating from a similar developmental trajectory such as the interconversion of two immune cell types, B cells (lymphoid cells) to macrophages (myeloid cells),<sup>9</sup> exocrine to endocrine pancreatic cells,<sup>10</sup> glial cells to neurons,<sup>11</sup> brown fat to muscle cells<sup>12</sup> and fibroblasts to cardiomyocytes.<sup>13</sup> All of these transdifferentiation events are within the same germ layer (i.e. mesoderm to mesoderm or neurectoderm to neurectoderm). Eventually, conversions between cells originating from different germ layers could also be achieved as fibroblast as well as hepatocytes could be directly converted into functional neurons with defined TF cocktails.<sup>14–16</sup> Apparently, even pronounced reprogramming barriers separating very distant cellular states can be crossed with small sets of lineage specifying TF proteins. The appreciation of this astonishing developmental plasticity sparked a plethora of studies attempting to re-direct cell fates. For example, blood cells,<sup>17</sup> endothelial cells,<sup>18</sup> hepatocytes,<sup>19–21</sup> sertoli cells<sup>22</sup> and thymic epithelial cells<sup>23</sup> could be successfully generated with TF cocktails. Readers are referred to excellent reviews that discuss the progress of cellular reprogramming and lineage conversions in more detail.<sup>24–33</sup> Remarkably, particularly potent reprogramming factors such as Oct4 were found to be able to

induce reprogramming alone in certain cell types and other cell types with small molecule supplementation.<sup>34–36</sup> Moreover, a few studies have reported that TFs can be omitted altogether as induced pluripotent or neuronal cells could be produced solely with small molecules.<sup>37,38</sup> However, efficiency of chemical reprogramming is considerably lower than TF based approaches and applications currently remain limited to mice. Hence, TF based cell lineage conversions continue to be the most effective and versatile approach. Therefore, we will focus our discussion on efforts to enhance TF based cell lineage conversions through protein engineering.

### ROADBLOCKS ON THE WAY TO FUNCTIONAL CELLS

The excitement sparked by cellular reprogramming is catalyzed by its promise to lead to new clinical applications. One strategy is to conduct “*in vitro* clinical trials.”<sup>26,39</sup> That is, cells obtained from patients through biopsies, blood or urine samples are differentiated into disease-relevant cell types. Next, preselected drugs or drug libraries can be assessed for their toxicity and potential to exert curative effects on those cells. It is hoped that this approach will accelerate personalized therapies, facilitate drug discovery and avoid the prescription of drugs that are toxic or ineffective to certain patient populations. Moreover, reprogramming technologies can be used to model human diseases in a dish. Here, the behavior of cells derived from patients is compared to cells from healthy donors. If disease-causing mutations are known, the mutation can be engineered using genome editing technologies and genetically matched isogenic cell lines can be studied. This way, diseases can be understood at an unprecedented depth, cellular pathways can be mapped, biomarkers can be discovered and therapeutic strategies can be developed. Lastly, the holy grail of stem cell research is to produce functional cells that can be transplanted back into patients to remedy degenerative diseases.<sup>40</sup> Encouragingly, diseases could be cured through cell therapies in animal models. For example, gene-corrected iPSC derived hematopoietic progenitors transplanted back into humanized sickle cell anemia mouse models could cure the animals.<sup>41</sup> This has led to the hope that diseases caused by deficiencies in well-defined cell types such as type 1 diabetes,<sup>42</sup> Parkinson’s disease<sup>43</sup> and retinal degeneration<sup>44</sup> are curable with cell-based therapies. Though, hematopoietic stem cells have been used in bone marrow transplants since the 1950’s, cell therapies in humans still pose major challenges, and daunting roadblocks remain. Most importantly, safety has to be rigorously assessed before transplanting the reprogrammed cells. iPSCs resemble cancer cells in many ways and are teratogenic when injected into mice. This poses a significant risk as incomplete differentiation, and remnant pluripotent cells could potentially lead to cancerous growth.<sup>45,46</sup> Collectively, avoiding insertional mutagenesis, oncogenic TFs and pluripotent reprogramming intermediates could solve this problem. Furthermore, it is often problematic to terminally differentiate cells so that they fully replicate the function of the cells matured *in vivo*. Cells have to be stable and need to be expandable so that they can be produced in sufficient quantities needed to support transplantation medicine. Ideally, reprogramming strategies should leave the genome unscathed, utilize cells that are genetically matched to the recipient with just the disease-causing loci corrected and produce an epigenetic state identical to the tissue embedded cell they are meant to replace. While optimized factor cocktails, novel culture conditions and small molecule compounds will likely further advance reprogrammed cells toward the clinic, we surmise that the engineering of the reprogramming TFs themselves provides a viable strategy to be further explored.

### DESIGNING BETTER PROTEINS

Bioengineering proteins to either enhance their activity or to install completely novel functions has been successfully accomplished in numerous instances. Day-to-day laboratory operations utilize a range of artificially enhanced proteins. Those include DNA polymerases with thoroughly optimized fidelity,<sup>47–49</sup> proteases with engineered activity and substrate specificity<sup>50</sup> and fluorescent proteins with increased brightness.<sup>51,52</sup> Likewise, protein therapeutics are often rationally improved. In particular coagulation factors to treat bleeding disorders were bioengineered in a variety of ways.<sup>53</sup> For example, factor IX was engineered to have prolonged activity by fusing it to a Fc fragment<sup>54</sup> and the coagulation factor VIIa was rationally mutagenized for inactivation resistance and optimized secretion profiles.<sup>55,56</sup> More ambitious goals include the engineering of whole pathways leading to the biotechnological synthesis of new products.<sup>57,58</sup>

What are the methods protein designers use to achieve their engineering goals? A rather simple way is to concatenate functional protein domains or even whole proteins. Examples include fusions of green fluorescent protein with antibody fragments that increase their brightness<sup>59</sup> or attaching effector domains such as nucleases to artificial TFs with customized DNA sequence preferences.<sup>60</sup> In addition, functional regions such as phosphorylation sites, protease cleavage sites, and signaling sequences can be rationally modified to install desired properties.<sup>55,56</sup> Most commonly, rational and randomization strategies are combined to achieve the desired results. Using the knowledge of the protein’s structure, sequence conservation and functional insights gained from site-directed mutagenesis experiments can lead to the selection of functionally important structural elements. Such elements could be individual or a small set of amino acids, secondary structure elements or subdomains. Frequently, design efforts target catalytic centers, substrate binding pockets or macromolecular contact interfaces. Those elements can then be modified taking biophysical parameters such as charge, size, and hydrophobicity, as well as functional data and sequence information of homologs into account. Yet, rationally predicting how a specific structural modification affects protein activity is a daunting task as our understanding about the structural basis for protein function remains limited. Therefore, protein designers often subject, structural elements earmarked for protein optimization to directed evolution.<sup>61</sup> This strategy requires a carefully designed randomization strategy, which can include error-prone polymerase chain reaction,<sup>62</sup> site-directed mutagenesis with randomized oligos<sup>63</sup> and “chimeragenesis,” that is the recombination of protein fragment libraries.<sup>64</sup> Libraries of modified proteins now undergo a screening and selection procedure to identify variants with improved functionality. Selection systems include binding assays such as phage display,<sup>65</sup> ribosome display,<sup>66</sup> enzymatic assays,<sup>67</sup> tests for protein stability,<sup>68</sup> genetic complementation combined with phenotypic read-outs<sup>69</sup> and *in vitro* compartmentalization.<sup>70</sup> Obviously, selection system design is critical as desired protein variants would escape detection if the screen cannot rigorously discriminate between enhanced and unwanted variants of the designed protein.<sup>61</sup>

Remarkably, efforts are being made to design proteins entirely from scratch using fragment libraries of nonnatural peptide sequences with minimal architectural constraints. Given the mindboggling number of theoretically possible protein sequences this seems like a herculean feat. Nevertheless, *de novo* design has led to the creation of some functional sequences.<sup>71,72</sup> Thus far, examples for the engineering of TF proteins are still rather rare. Here we ask whether the toolkit of protein engineering could be employed to design reprogramming TFs to more

effectively engineer cell lineage conversions and to bring progress to regenerative biomedicine.

## ENGINEERING SYNTHETIC REPROGRAMMING FACTORS

### Enhancing reprogramming efficiency with potent transactivation domains

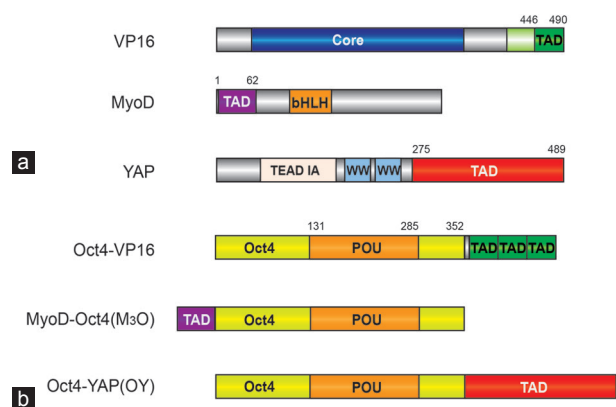
The optimization of reprogramming strategies has been a priority for many laboratories as the original protocol was rather inefficient. Efficiency enhancements could be achieved by supplementing the media,<sup>73</sup> altering the factor cocktails,<sup>74–77</sup> changing the sequence of factor addition,<sup>78</sup> adding small molecules<sup>35,79–84</sup> or removing reprogramming roadblocks.<sup>85,86</sup> In addition, some studies resorted to protein engineering to improve reprogramming (Table 1). Based on the assumption that reprogramming TFs mainly act by inducing mRNA synthesis of their target genes, several engineering efforts were made to increase the transactivation potential of TFs by fusing them to potent transactivation domains (TADs) (TAD-TFs, Figure 1).

### Viral protein 16-transactivation domain

The viral protein 16 (VP16) is a 490 amino acids TF protein of the herpes simplex virus with strong transactivating activity (Figure 1a). Its potent TAD was mapped and molecularly dissected more than 25 years ago and found to consist of an acidic C-terminal region spanning approximately 80 amino acids.<sup>87,88</sup> Immediately after its discovery the VP16-TAD has been utilized to engineer chimeric TFs with enhanced activity.<sup>89</sup> More recently, VP16 has also been utilized to enable cellular reprogramming. When the VP16-TAD was fused to the *Xenopus* ortholog of the pancreatic and duodenal homeobox 1 (Pdx1), a chimeric protein could induce the conversion of liver cells to pancreatic cell in transgenic tadpoles.<sup>90</sup> A similar Pdx1-VP16 fusion induced insulin biosynthesis and ameliorated glucose tolerance in mouse diabetic models.<sup>91,92</sup> In an effort to enhance iPSC formation the VP16-TAD was fused to pluripotency reprogramming factors.<sup>93</sup> In this study, a core fragment of the VP16-TAD (residues 446–490) was attached to Oct4, Sox2, Klf4 and Nanog TFs separated by a glycine-rich linker (Figure 1a). With the exception of Klf4, the engineered TAD-TFs

substantially outperformed the wild-type proteins with regards to both the efficiency and the kinetics of iPSC generation in mouse and human cells. Moreover, Oct4-VP16 alone could efficiently reprogram mouse embryonic fibroblasts into germline-competent iPSCs.<sup>93</sup> An Oct4 construct containing three C-terminal VP16 copies arranged in tandem exhibited the highest efficiency (Figure 1b).

A separate study also reported that fusions of VP16 to mouse Oct4, human Oct4 and *Xenopus* Xlpou91 could support reprogramming as well as rescue Oct4 null ESCs.<sup>94</sup> However, the authors did not observe a substantial enhancement in the reprogramming efficiency by the engineered proteins. The differences between the two studies could be caused by variations in the reprogramming conditions and construct



**Figure 1:** Enhancing reprogramming efficiency with TAD-TF chimeras. (a) Domain structure of VP16, MyoD and YAP from which TADs were derived. (b) Domain structure of chimeric Oct4-TAD proteins demonstrated to potentially enhance the reprogramming of somatic cells to iPSCs.<sup>93,97,103</sup> YAP1 was drawn according to Uniprot-ID P46938 and VP16 and MyoD according to Hirai *et al.*,<sup>99</sup> 2010. TAD: transactivation domain; bHLH: basic helix-loop-helix domain; POU: Pit1-Oct-Unc-86 related domain; TEAD-IA: interaction region with the TF TEAD; TFs: transcription factors; Oct4: octamer binding protein 4; iPSCs: induce pluripotent stem cells; VP16: viral protein 16; YAP: yes-associated protein.

**Table 1: Engineered reprogramming factors**

Protein modified	Type of modification	Effects on reprogramming and differentiation	References
Reprogramming factor modulation with potent TADs			
Oct4	Fusing the MyoD TAD to Oct4	~50-times more iPSC colonies compared to OSKM	97
Oct4, Sox2, Nanog	VP16-TAD linked to Oct4, Sox2 and Nanog	Enhanced reprogramming of mouse and human fibroblasts; single factor reprogramming with Oct4-VP16	93
Oct4 and Xlpou91	Vp16 linked to Oct4 and its <i>Xenopus</i> orthologue	iPSC induction similar to WT cocktail	94
Sox2, Sox18, Sox4	Sox17-TAD linked to Sox2, Sox18EK and Sox4EK	~3–5 more iPSC colonies than Sox2-WT	95
Oct4, Sox2, and Nanog	YAP-TAD linked to Oct4, Sox2 and Nanog	~100-fold iPSC induction efficiency increase and 24 h reprogramming	103
Pdx1	VP16-TAD fused to <i>Xenopus</i> Pdx1	Conversion of liver to pancreas in <i>Xenopus</i> tadpoles	90
Pdx1	VP16-TAD fused to mouse Pdx1	Induces insulin synthesis and ameliorates glucose tolerance in diabetic mouse models	91,92
MyoD	DNA-binding domain of MyoD fused to the TAD of myocardin	Improves directed differentiation of human mesenchymal stem cells into skeletal myocytes	149
Mef2c	MyoD TAD linked to Mef2c	15-fold efficiency increase in fibroblasts to cardiomyocytes conversion	102
Targeted activation of endogenous reprogramming factors			
TALEs	Designer TALEs with VP64 targeting the distal Oct4 enhancer	Substitutes for exogenous Oct4 to generate iPSCs with comparable efficiency	120
Reprogramming factors with engineered DNA recognition properties			
MyoD and E12	Swap of three amino acids from MyoD into E12	Mutant E12 can convert fibroblasts into myocytes	122
Sox17 and Sox2	Point mutation in the DNA binding domain of Sox17	Conversion of Sox17 into an iPSC inducer with about 4-fold greater efficiency than Sox2	142

TAD: transactivation domain; Oct4: octamer binding protein 4; iPSC: induce pluripotent stem cells; MEFs: mouse embryonic fibroblast; TALEs: transcription-activator-like effectors; YAP: yes-associated protein; Pdx1: pancreatic and duodenal homeobox 1; OSKM: Oct4, Sox2, Klf4, and c-Myc; VP16: viral protein 16; Klf4: krüppel-like factor 4; WT: wild-type

design as different VP16 fragments, linkers, and VP16 copy numbers were used. The potency of the VP16-TAD for iPSC generation was again highlighted when it could be shown that fusions of VP16 to the truncated DNA binding high-mobility group (HMG) domain of Sox2 imparted some reprogramming activity to the otherwise inactive HMG fragments.<sup>95</sup>

#### **MyoD-transactivation domain**

Inspired by the remarkable potency of MyoD to single-handedly convert fibroblasts into muscle cells<sup>96</sup> Hirai *et al.*<sup>97</sup> asked whether the TAD of MyoD can enhance reprogramming to pluripotency. MyoD, a TF of the basic helix-loop-helix (bHLH) family,<sup>98</sup> contains a ~60 amino acid TAD at its N-terminus<sup>99</sup> (**Figure 1a**). The authors generated a series of MyoD TAD fragments and generated chimeric proteins by attaching them to the full-length mouse Oct4 protein. A chimeric protein consisting of a MyoD fragment (residues 1–62) attached to the N-terminus of Oct4 (termed M<sub>3</sub>O, **Figure 1b**) was found to strongly amplify the number of iPSC colonies in reprogramming assays using mouse and human cells. However, increasing the copy number of the M<sub>3</sub> fragment was detrimental to iPSC generation. Notably, neither Klf4 nor Sox2 could be enhanced with the M<sub>3</sub> fragment. Rather M<sub>3</sub>Sox2 and M<sub>3</sub>Klf4 fusions inhibited reprogramming. Likewise, replacing M<sub>3</sub> with the TADs of Gata4, Mef2c, Tax and Tat eliminated the reprogramming activity of Oct4. Collectively, the engineered M<sub>3</sub>O factor appears to support reprogramming in a more specific and context-dependent manner in contrast to the VP16 fusions that allow a more flexible design.<sup>97</sup> It was subsequently found that when culturing the cells in serum-free media at low density, M<sub>3</sub>O containing cocktails could enhance iPSC formation to 26% and 7% of the transfected mouse or human cells, respectively.<sup>100</sup> Moreover, when tested side-by-side, the M<sub>3</sub>O construct outperformed Oct4-VP16 fusions.<sup>100</sup> Consistently, in a study done by another group, M<sub>3</sub>O was found to accelerate reprogramming based on a cocktail using modified messenger RNAs.<sup>101</sup> Motivated by the success in using the M<sub>3</sub> domain to optimize Oct4 activity, Hirai *et al.*<sup>102</sup> went on to ask whether factors involved in cardiac transdifferentiation can be improved through M<sub>3</sub> chimeras. While M<sub>3</sub> fusions to Gata4, Tbx5 and Hand2 had either no or adverse consequences, M<sub>3</sub>-Mef2c chimeras lead to a 15-fold increase in the number of beating clusters of induced cardiomyocytes (iCMs). Mef2c-VP16 fusions also showed some increase in iCM formation, albeit to only 20% of the M3Mef2c levels.

#### **Yes-associated protein-transactivation domain**

More recently, the TAD of the yes-associated protein (YAP) was used to engineer iPSC inducing TFs (**Figure 1a**).<sup>103</sup> YAP is a downstream effector of the Hippo signaling pathway and co-activates transcription in concert with TFs of the TEAD family by recruiting histone methyltransferases.<sup>104</sup> YAP promotes oncogenesis as well as the self-renewal of ESCs via its potent transactivation activity.<sup>105,106</sup> Its C-terminal TAD was found to activate reporter genes as potently as the VP16 TAD.<sup>107</sup> Zhu *et al.*<sup>103</sup> fused the C-terminal YAP residues 275–489 to the C-termini of Oct4, Sox2 and Nanog to generate the engineered proteins Oy, Sy and Ny (**Figure 1b**). When this cocktail was used for iPSC induction in combination with native Klf4, the reprogramming efficiency rose from <1% to ~40%. Furthermore, the reprogramming kinetics was markedly accelerated with iPSC colonies appearing on the day after switching the transfected fibroblasts to ESC medium. The Sox2-YAP fusion was reported to be most critical for the acceleration among the three modified TFs.

What do the three TADs used to engineer reprogramming TFs have in common? They consist of predominantly acidic and hydrophobic

residues leading to a strongly negative net charge. However, none of the TADs has been structurally characterized presumably due to their flexible structure in the absence of binding partners. Therefore, the molecular details of how those TADs create a chromatin environment instructive for the mRNA synthesis of nearby genes remains unclear. While a series of TAD interaction partners were previously detected,<sup>99,108</sup> the affinity and selectivity of those TADs for co-regulators such as p300 or the mediator complex has not yet been studied in a systematic manner. Hence, whether those TADs mediate a general, or a TAD specific mechanism of transcriptional activation awaits further exploration. Collectively, there appear to be no obvious rules of how TAD-TF chimeras should be designed to engineer lineage converting TFs. Rather, optimal constructs had to be empirically produced for each TAD-TF combination.<sup>93,94,97</sup> For example, increasing TAD copy numbers can either boost<sup>93</sup> or impede TAD-TF activity.<sup>94,97,109</sup> Further parameters to be optimized include the length of the TAD fragment used, the position of the TADs at either the N- or C-termini of the TFs, and the inclusion of linker sequences.

#### **Inducing endogenous reprogramming factors with TAL effectors**

Artificial proteins based on C2-H2 zinc-finger proteins (ZFPs), transcription activator-like effectors (TALEs) and RNA-guided clustered regularly interspaced short palindromic repeat (CRISPR) Cas (CRISPR associated) can be designed to target genomic loci with high specificity. Typically, those proteins are constructed to contain nuclease effector domains that enable genome editing at single base-pair resolution (reviewed in<sup>110,111</sup>). While off-target effects had been a lurking concern, whole genome sequencing studies demonstrated that unwanted modifications are very rare.<sup>112,113</sup> Recently, designer TALEs (dTALEs) and ZFPs have also been used to engineer transcriptional activators.<sup>114,115</sup> TALEs consist 33–34 amino acid repeat domains with hypervariable residues at position 12 and 13.<sup>116</sup> The identity of the dipeptide at positions 12/13 determines a recognition code (HD = C, NG = T, NI = A, NS = A, C, G or T, NN = A or G), which allows to rationally create dTALEs that recognize DNA sequences of choice.<sup>116</sup> As up to 33 TALE repeat domains can be arranged in tandem, genomic loci can be targeted with high precision. Although, dTALE design is somewhat more straightforward than ZFP design, initial efforts were undertaken using ZFP-VP16 fusions constructed to target a sequence proximal to the transcriptional start site of Oct4.<sup>117</sup> In another attempt, the fusion of a KRAB domain fused to a designer ZFP could activate endogenous Oct4 protein in series of cell lines.<sup>118</sup> This was a surprising observation because the KRAB conventionally acts as transcriptional repressor. dTALE-VP16 fusions designed to bind proximal promoter sequences of SOX2, Klf4, c-MYC and Oct4 could activate reporter constructs, but only dTALEs targeting Klf4 and SOX2 could also activate the endogenous genes in 293FT cells.<sup>60</sup> Moreover, dTALEs targeting the proximal promoter of Oct4 could activate the gene in NSCs where it is otherwise silenced. However, this strategy required the addition of histone deacetylase or DNA methyltransferase inhibitors suggesting that some chromatin loosening is needed for the dTALE-TF to access its target site.<sup>119</sup>

Gao *et al.*<sup>120</sup> asked whether dTALE-TFs can replace conventionally used reprogramming TFs by activating their endogenous counterparts. The authors used VP64-TADs (four tandem repeats of VP16) to engineer designer transcription activators (A-dTF) that target distal enhancers of reprogramming factors.<sup>120</sup> Indeed, a A-dTF designed to activate endogenous Oct4 could replace exogenous Oct4 and induce iPSCs in combination with c-Myc, Klf4 and Sox2. While reprogrammed cells appeared faster when using the A-dTF the overall iPSC colony yield

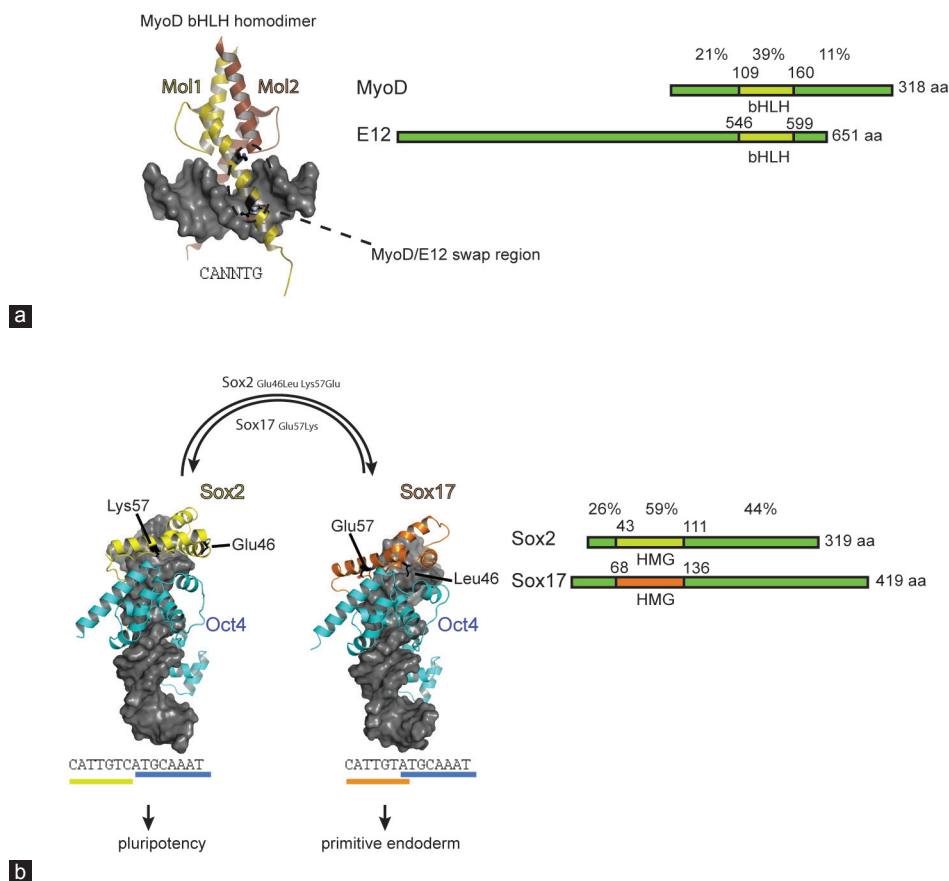
was higher when Oct4 was used directly.<sup>120</sup> The authors went on to show that A-dTF, targeting a distal Nanog enhancer could convert epiblast stem cells into ESCs. This study provides an elegant proof-of-concept that TALE-based TFs can replace native reprogramming factors. However, as the sole function of dTALE-TFs is to activate endogenous reprogramming factors that would eventually have to finish the job, it remains to be demonstrated whether this method can enhance cell lineage conversions.

### Engineering chromatin association of reprogramming factors

#### Turning E12 into a myogenic transcription factor

The TAD-TFs and TFs endogenously activated by A-dTFs will likely engage the genome in the same manner as the native TFs as the DNA recognition domain is not modified. So far, only a few engineering efforts focused on protein interfaces involved in DNA recognition that would alter their genomic binding profile. Still, several swap experiments that install new functions and create engineered reprogramming TFs have been successful. The reprogramming pioneer Weintraub had provided the first evidence that lineage conversion activity of reprogramming TFs can be radically interchanged with strategically placed point mutations at the DNA contact interface.<sup>121,122</sup> Following the seminal discovery that

MyoD alone can induce a myogenic program in fibroblasts,<sup>3</sup> Weintraub *et al.* continued to dissect the molecular basis of its specific activity. MyoD belongs to the bHLH family of TFs whose members bind to short palindromic CANNTG E-box motifs as homo- or heterodimers. By adopting a scissor-like architecture, bHLH TFs bind the major groove of the DNA through the basic regions of helix1<sup>98</sup> (Figure 2a). The E-box can be bound by most bHLH with similar affinity and many amino acids contacting the DNA are highly conserved. Nevertheless, some subtly different sequence preferences have recently been detected which could contribute to distinctive roles of bHLH TFs in cell fate determinations.<sup>123</sup> Weintraub compared the DNA recognition and the reprogramming potential of MyoD and E12, another bHLH factor that is ubiquitously expressed and does not trigger myogenesis.<sup>121,122</sup> By grafting just three amino acids from MyoD into E12: N114A and N115T of the basic region and D124K of the linker (MyoD numbering; residues 6, 7 and 17 – numbering according to current bHLH conventions), E12 acquired the ability to convert fibroblasts into muscle cells<sup>121,122</sup> (Figure 2a). As the Ala-Thr dipeptide is conserved in myogenic bHLH TFs, this sequence is critical for executing a myogenic gene expression program. Surprisingly, the large degree of sequence variation between MyoD and E12 outside the bHLH domain did not contribute to their cell type



**Figure 2:** Rational engineering how reprogramming factors read genomes. **(a)** Dimeric MyoD structure bound to DNA (pdb-ID 1 mdy<sup>98</sup>). The DNA is shown as a gray surface plot and the two MyoD molecules forming the dimer are shown as yellow and brown cartoon. Residues engrafted from MyoD into E12 to turn E12 into a myogenic protein are shown as black ball-and-sticks and are highlighted with a dashed oval. **(b)** Structural models showing heterodimers of Oct4 (light blue) and Sox2 (yellow) on the canonical motif and Oct4-Sox17 (orange) on the compressed motif. Sox residues that determine the discriminative heterodimerization with Oct4 on canonical and compressed motifs are shown as black ball-and-sticks. Transplanting Lys57 from Sox2 into Sox17 alone turns Sox17 into a potent inducer of pluripotency.<sup>142</sup> The structural models were constructed as described in<sup>146</sup> based on coordinates derived from pdb-IDs 3f27 and 1 gt0.<sup>129,130,145</sup> To the right of the structural cartoons the domain structure of MyoD versus E12 and Sox2 versus Sox17 is compared. The percentages above the domain plots indicate the amino acid identity between the protein pairs in the N-terminal, DNA binding and C-terminal region. The alignment was performed using sequences derived from Uniprot: MyoD:P10085; E12:P15806; Sox2:P48432; Sox17:Q61473. Oct4: octamer binding protein 4.

specific functions (**Figure 2a**). Rather, just three amino acids at the DNA interaction surface specify their functional diversity. Subsequent studies suggested that the Ala-Thr dipeptide affects the conformation of Arg-111 in the basic region and thereby modulates the access of Arg-111 to the major groove of the DNA binding site.<sup>124</sup> Those rearrangements at the DNA-binding interface could translate into allosteric events at other interfaces, such as binding sites for chromatin modifiers, and thereby influence the functional consequences of the binding event.<sup>125</sup>

### Turning Sox17 into a pluripotency inducer

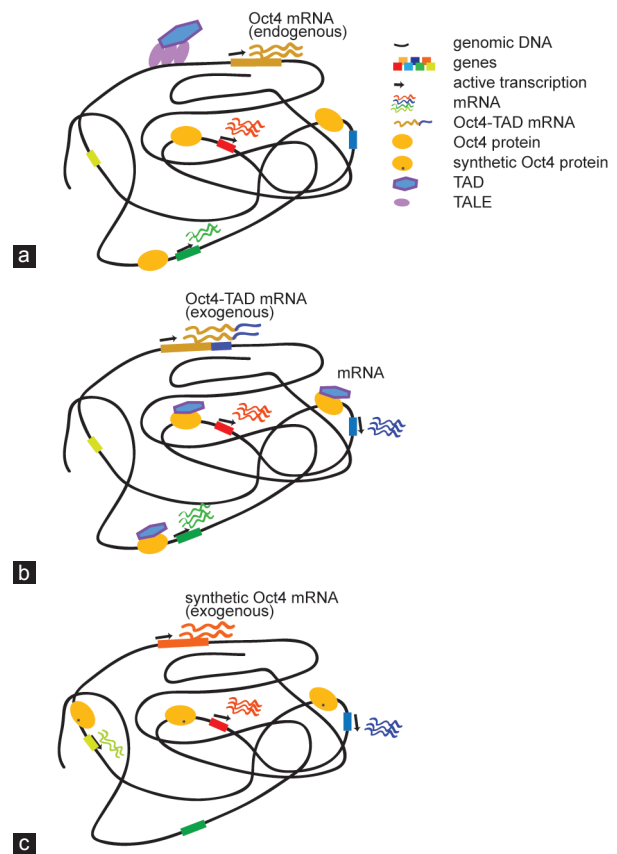
Our laboratory has made efforts to scrutinize the mechanism how proteins of the 20-member Sox family recognize their DNA target sites. Confusingly, all Sox proteins bind a near-identical CATTGT-like sequence<sup>126,127</sup> and engage DNA by binding to the minor groove to induce a 70° kink using a conserved set of amino acids.<sup>128–136</sup> How then can individual members select specific gene-sets to initiate characteristic cell fate decisions? The DNA binding HMG domain of Sox protein not only mediates sequence-specific DNA recognition but is also the main determinant of a partner code enabling selective interactions with other TFs.<sup>137–141</sup> By conducting quantitative electrophoretic mobility shift assays to study the HMG mediated partnership with the Pit1-Oct-Unc-86 (POU) domain of Oct4, we observed different propensities of Sox-family members to heterodimerize with Oct4 on a series of differently configured composite sox-oct binding sites.<sup>133,142</sup> In particular an unusual “compressed” element—where one nucleotide separating the sox and oct half sites is removed – still recruits Sox17/Oct4 dimers, whereas Sox2/Oct4 dimers can no longer assemble (**Figure 2b**). Conversely, the Sox2/Oct4 pair dimerizes markedly better on the canonical motif than the Sox17/Oct4 pair. In the search for the structural basis for these differences, a single amino-acid at position 57 of the HMG caught our attention. This residue, a Lys in Sox2 and a Glu in Sox17, shows a high degree of sequence variation amongst paralogous Sox proteins although it occupies a critical position at the Oct4 interaction interface.<sup>129,130,132,133</sup> By exchanging this residue between Sox2 and Sox17 to produce Sox17EK and Sox2KE proteins, highly cooperative dimer formation of the Sox17EK/Oct4 complex on the canonical motif is installed. The wild type Sox2 normally partners with Oct4 in OKSM<sup>4,5</sup> or OSNL (OS plus Nanog and Lin28)<sup>6</sup> cocktails to activate the pluripotency circuitry. By contrast, the wild-type Sox17 induces endoderm differentiation when overexpressed in ESCs. The activity of the engineered factors was, therefore, studied in iPSC generation assays.<sup>142,143</sup> When we replaced Sox2 with Sox17EK in OSKM cocktails, we could induce iPSCs with improved efficiency in both mouse<sup>142</sup> and human cells.<sup>95</sup> An analogously modified Sox7EK protein showed a similar behavior, whereas Sox4 and Sox18 needed additional TAD engineering for their conversion into reprogramming TFs.<sup>95</sup> Using chromatin immunoprecipitation followed by high-throughput sequencing, we found that Sox17EK and Sox2 show a very similar binding profile when overexpressed in mouse ESCs.<sup>144</sup> Both proteins pair with Oct4 on many genomic loci earmarked with canonical sox-oct motifs. By contrast, Sox17 partners with Oct4 on enhancers containing the compressed motif. Apparently, a single point mutation drastically changed how Sox proteins co-select their target genes by partnering with Oct4. Yet, the converse Sox2KE mutant could neither effectively dimerize with Oct4 on the canonical nor on the compressed sequence. This puzzle was resolved more recently when a novel Oct4 crystal structure and molecular dynamics simulation suggested an additional discriminatory interaction between residue 46 of Sox proteins with an Oct4 specific helix in the POU linker.<sup>145,146</sup> Indeed, a rationally designed Sox2E46LK57E double mutant now cooperatively dimerizes with Oct4

on the compressed motif. It will be of interest to explore the activity of this engineered Sox factor in lineage conversion experiments.

Collectively, the MyoD and Sox17EK examples show that the cell fate conversion potential of reprogramming TFs can be drastically changed with rather minimal modifications at structurally critical interfaces. We surmise that these insights could be utilized to engineer more potent and safer reprogramming TFs. Contrary to the TAD-TF and the TALE-TF approach; TFs with engineered DNA-binding domain likely engage the genome in a new manner (**Figure 3**). This way, it could be possible to break reprogramming barriers more effectively and to direct cells trapped in a local minimum of the Waddington landscape towards a desired state of differentiation.

### OUTLOOK — NOVEL APPROACHES FOR REPROGRAMMING FACTOR DESIGN

To produce cells for clinical applications, the process should be tightly controlled, fast and exclude undesired by-products. In particular,



**Figure 3:** Regulatory outcome of different categories of engineered reprogramming TFs. (a) TALEs coupled with TADs can be engineered to switch on otherwise silenced reprogramming TFs. By targeting the distal enhancer of Oct4 its expression is activated. The resulting gene expression program is expected to resemble the wild-type scenario. (b) When potent TADs are fused to reprogramming TFs the chimeric protein is expected to target genomic regions reminiscent of the unmodified wild-type protein (a). However, the presence of a potent TAD elevates expression levels as compared to the wild-type and may also trigger the activation of genes that would otherwise be silent. (c) Rationally placed point mutations within the DNA binding domain can modify DNA recognition principles and alter the genomic binding profile. The cartoon represents a synthetic Oct4 with modified binding preferences that leads to the activation of genes that the wild-type protein would not switch on. TFs: transcription factors; TALEs: transcription activator-like effectors; TAD: transactivation domain; Oct4: octamer binding protein 4.

reprogramming to pluripotency has witnessed a multitude of studies aimed to improve the efficiency of iPSC generation (excellent reviews by Papp and Plath<sup>28</sup> and Soufi<sup>31</sup>). Initially, iPSC generation was rather slow and only a small number of cells transfected with a cocktail of reprogramming TFs could be reprogrammed.<sup>4</sup> Confusingly, it appeared that there is a high degree of randomness in cell populations and by simple chance a small subset of cells enters a path leading to the successive progression towards pluripotency in a more deterministic fashion.<sup>147,148</sup> Yet, as roadblocks toward the pluripotency continue to be removed; fully controlled and efficient iPSC generation could soon be achieved.<sup>85,86</sup> As the quality of iPSCs produced by engineered reprogramming factors was validated by examining their contribution to embryonic development and the capacity for germline transmission, synthetic TFs could still contribute to the ultimate cocktail.<sup>93,97,100,103,120</sup> However, iPSCs are only an intermediary by-product on the way towards transplantation-grade functional cells. To lower the risk of cancerogenesis, a pluripotent intermediate should be avoided altogether or, minimally, complete differentiation of formerly pluripotent cells has to be ensured. Reproducibly generating functional cells to cure degenerative diseases will remain a challenge in the years to come. We anticipate that protein engineering techniques can help to overcome reprogramming barriers and better control cell lineage conversions to produce functional cells more safely and with properties more closely matching their *in vivo* counterparts (Figure 3). While widely used in fields such as enzymology and immunology, protein engineering is still in its infancy in cellular reprogramming. This is partly because of our incomplete understanding of how TFs work. Our structural knowledge is mostly restricted to isolated domains bound to short stretches of DNA. The mechanism of DNA target site selection, chromatin opening and how TFs stimulate mRNA synthesis remains largely unclear. Nevertheless, the studies highlighted in this review testify the promise of the approach and warrant further exploration as to whether protein engineering can bring stem cell biology closer to the bedside.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## ACKNOWLEDGMENTS

Our laboratory is supported by the people's government of Guangzhou municipality Science and Technology Project 2011Y2-00026 and by a 2013 MOST China-EU Science and Technology Cooperation Program, Grant No. 2013DFE33080. V.M. thanks the CAS-TWAS President's Fellowship and UCAS for financial and infrastructure support. We thank Andrew Hutchins, Jiekai Chen, Ajaybabu Pobbati, Swaine Chen and Jörg Kahle for insightful comments.

## REFERENCES

- Haeckel EH. Zur Entwicklungsgeschichte der Siphonophoren. Utrecht: C. Van Der Post Jr; 1869.
- Gurdon JB. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* 1962; 10: 622–40.
- Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 1987; 51: 987–1000.
- Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; 126: 663–76.
- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, *et al*. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; 131: 861–72.
- Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, *et al*. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; 318: 1917–20.
- Cai J, Zhang Y, Liu P, Chen S, Wu X, *et al*. Generation of tooth-like structures from integration-free human urine induced pluripotent stem cells. *Cell Regen (Lond)* 2013; 2: 6.
- Waddington CH. Organisers and Genes. The University Press; 1940.
- Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages.

- Cell* 2004; 117: 663–76.
- Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. *In vivo* reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 2008; 455: 627–32.
- Heins N, Malatesta P, Ceconi F, Nakafuku M, Tucker KL, *et al*. Glial cells generate neurons: the role of the transcription factor Pax6. *Nat Neurosci* 2002; 5: 308–15.
- Seale P, Bjork B, Yang W, Kajimura S, Chin S, *et al*. PRDM16 controls a brown fat/skeletal muscle switch. *Nature* 2008; 454: 961–7.
- Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, *et al*. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 2010; 142: 375–86.
- Pang ZP, Yang N, Vierbuchen T, Ostermeier A, Fuentes DR, *et al*. Induction of human neuronal cells by defined transcription factors. *Nature* 2011; 476: 220–3.
- Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, *et al*. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 2010; 463: 1035–41.
- Marro S, Pang ZP, Yang N, Tsai MC, Qu K, *et al*. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* 2011; 9: 374–82.
- Szabo E, Rampalli S, Risueño RM, Scherch A, Mitchell R, *et al*. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* 2010; 468: 521–6.
- Margariti A, Winkler B, Karamariti E, Zampetaki A, Tsai TN, *et al*. Direct reprogramming of fibroblasts into endothelial cells capable of angiogenesis and reendothelialization in tissue-engineered vessels. *Proc Natl Acad Sci U S A* 2012; 109: 13793–8.
- Huang P, He Z, Ji S, Sun H, Xiang D, *et al*. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 2011; 475: 386–9.
- Huang P, Zhang L, Gao Y, He Z, Yao D, *et al*. Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell* 2014; 14: 370–84.
- Sekiya S, Suzuki A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* 2011; 475: 390–3.
- Buganim Y, Itskovich E, Hu YC, Cheng AW, Ganz K, *et al*. Direct reprogramming of fibroblasts into embryonic Sertoli-like cells by defined factors. *Cell Stem Cell* 2012; 11: 373–86.
- Bredenkamp N, Ulyanchenko S, O'Neill KE, Manley NR, Vaidya HJ, *et al*. An organized and functional thymus generated from FOXP1-reprogrammed fibroblasts. *Nat Cell Biol* 2014; 16: 902–8.
- Bellin M, Marchetto MC, Gage FH, Mummery CL. Induced pluripotent stem cells: the new patient? *Nat Rev Mol Cell Biol* 2012; 13: 713–26.
- Graf T. Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* 2011; 9: 504–16.
- Grskovic M, Javaherian A, Strulovici B, Daley GQ. Induced pluripotent stem cells – opportunities for disease modelling and drug discovery. *Nat Rev Drug Discov* 2011; 10: 915–29.
- Hanna JH, Saha K, Jaenisch R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 2010; 143: 508–25.
- Papp B, Plath K. Epigenetics of reprogramming to induced pluripotency. *Cell* 2013; 152: 1324–43.
- Sabour D, Schöler HR. Reprogramming and the mammalian germline: the Weismann barrier revisited. *Curr Opin Cell Biol* 2012; 24: 716–23.
- Sánchez Alvarado A, Yamanaka S. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* 2014; 157: 110–9.
- Soufi A. Mechanisms for enhancing cellular reprogramming. *Curr Opin Genet Dev* 2014; 25: 101–9.
- Takahashi K, Yamanaka S. Induced pluripotent stem cells in medicine and biology. *Development* 2013; 140: 2457–61.
- Vierbuchen T, Wernig M. Molecular roadblocks for cellular reprogramming. *Mol Cell* 2012; 47: 827–38.
- Kim JB, Greber B, Araúzo-Bravo MJ, Meyer J, Park KI, *et al*. Direct reprogramming of human neural stem cells by OCT4. *Nature* 2009; 461: 649–3.
- Wang H, Cao N, Spencer CI, Nie B, Ma T, *et al*. Small molecules enable cardiac reprogramming of mouse fibroblasts with a single factor, Oct4. *Cell Rep* 2014; 6: 951–60.
- Wu G, Schöler HR. Role of Oct4 in the early embryo development. *Cell Regen (Lond)* 2014; 3: 7.
- Cheng L, Hu W, Qiu B, Zhao J, Yu Y, *et al*. Generation of neural progenitor cells by chemical cocktails and hypoxia. *Cell Res* 2014; 24: 665–79.
- Hou P, Li Y, Zhang X, Liu C, Guan J, *et al*. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* 2013; 341: 651–4.
- Heilker R, Traub S, Reinhardt P, Schöler HR, Sternecker J. iPSC cell derived neuronal cells for drug discovery. *Trends Pharmacol Sci* 2014; 35: 510–9.
- Daley GQ. The promise and perils of stem cell therapeutics. *Cell Stem Cell* 2012; 10: 740–9.
- Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, *et al*. Treatment of sickle cell anemia mouse model with iPSCs generated from autologous skin. *Science* 2007; 318: 1920–3.
- Pagliuca FW, Melton DA. How to make a functional  $\beta$ -cell. *Development* 2013; 140: 2472–83.
- Buttery PC, Barker RA. Treating Parkinson's disease in the 21<sup>st</sup> century: can stem cell transplantation compete? *J Comp Neurol* 2014; 522: 2802–16.



- 44 Jin ZB, Takahashi M. Generation of retinal cells from pluripotent stem cells. *Prog Brain Res* 2012; 201: 171–81.
- 45 Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007; 448: 313–7.
- 46 Ohnishi K, Semi K, Yamamoto T, Shimizu M, Tanaka A, *et al*. Premature termination of reprogramming *in vivo* leads to cancer development through altered epigenetic regulation. *Cell* 2014; 156: 663–77.
- 47 Patel PH, Kawate H, Adman E, Ashbach M, Loeb LA. A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J Biol Chem* 2001; 276: 5044–51.
- 48 Summerer D, Rudinger NZ, Detmer I, Marx A. Enhanced fidelity in mismatch extension by DNA polymerase through directed combinatorial enzyme design. *Angew Chem Int Ed Engl* 2005; 44: 4712–5.
- 49 Leconte AM, Patel MP, Sass LE, McInerney P, Jarosz M, *et al*. Directed evolution of DNA polymerases for next-generation sequencing. *Angew Chem Int Ed Engl* 2010; 49: 5921–4.
- 50 Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, *et al*. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci U S A* 2013; 110: 7229–34.
- 51 Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 1996; 173: 33–8.
- 52 Heim R, Cubitt AB, Tsien RY. Improved green fluorescence. *Nature* 1995; 373: 663–4.
- 53 Fogarty PF. Biological rationale for new drugs in the bleeding disorders pipeline. *Hematology Am Soc Hematol Educ Program* 2011; 2011: 397–404.
- 54 Peters RT, Low SC, Kamphaus GD, Dumont JA, Amari JV, *et al*. Prolonged activity of factor IX as a monomeric Fc fusion protein. *Blood* 2010; 115: 2057–64.
- 55 Swaroop M, Moussalli M, Pipe SW, Kaufman RJ. Mutagenesis of a potential immunoglobulin-binding protein-binding site enhances secretion of coagulation factor VIII. *J Biol Chem* 1997; 272: 24121–4.
- 56 Pipe SW, Kaufman RJ. Characterization of a genetically engineered inactivation-resistant coagulation factor VIIIa. *Proc Natl Acad Sci U S A* 1997; 94: 11851–6.
- 57 Schiel-Bengelsdorf B, Dürre P. Pathway engineering and synthetic biology using acetogens. *FEBS Lett* 2012; 586: 2191–8.
- 58 Rohlin L, Oh MK, Liao JC. Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design. *Curr Opin Microbiol* 2001; 4: 330–5.
- 59 Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, *et al*. Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* 2010; 17: 133–8.
- 60 Zhang F, Cong L, Lodato S, Kosuri S, Church GM, *et al*. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 2011; 29: 149–53.
- 61 Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 2009; 10: 866–76.
- 62 Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 1992; 2: 28–33.
- 63 Irague R, Tarquis L, André I, Moulis C, Morel S, *et al*. Combinatorial engineering of dextranase specificity. *PLoS One* 2013; 8: e77837.
- 64 Hiraga K, Arnold FH. General method for sequence-independent site-directed chimeragenesis. *J Mol Biol* 2003; 330: 287–96.
- 65 Simon MD, Sato K, Weiss GA, Shokat KM. A phage display selection of engrailed homeodomain mutants and the importance of residue Q50. *Nucleic Acids Res* 2004; 32: 3623–31.
- 66 Hanes J, Plückthun A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A* 1997; 94: 4937–42.
- 67 Guo J, Gaj T, Barbas CF 3<sup>rd</sup>. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol* 2010; 400: 96–107.
- 68 Asial I, Cheng YX, Engman H, Dollhopf M, Wu B, *et al*. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat Commun* 2013; 4: 2901.
- 69 Suzuki M, Baskin D, Hood L, Loeb LA. Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites *in vivo* with the crystal structure. *Proc Natl Acad Sci U S A* 1996; 93: 9670–5.
- 70 Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998; 16: 652–6.
- 71 Smith BA, Hecht MH. Novel proteins: from fold to function. *Curr Opin Chem Biol* 2011; 15: 421–6.
- 72 Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH. *De novo* designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth. *PLoS One* 2011; 6: e15364.
- 73 Esteban MA, Wang T, Qin B, Yang J, Qin D, *et al*. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* 2010; 6: 71–9.
- 74 Feng B, Jiang J, Kraus P, Ng JH, Heng JC, *et al*. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* 2009; 11: 197–203.
- 75 Han J, Yuan P, Yang H, Zhang J, Soh BS, *et al*. Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* 2010; 463: 1096–100.
- 76 Heng JC, Feng B, Han J, Jiang J, Kraus P, *et al*. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 2010; 6: 167–74.
- 77 Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, *et al*. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* 2008; 10: 353–60.
- 78 Liu X, Sun H, Qi J, Wang L, He S, *et al*. Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nat Cell Biol* 2013; 15: 829–38.
- 79 Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, *et al*. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell* 2009; 5: 491–503.
- 80 Li K, Zhu S, Russ HA, Xu S, Xu T, *et al*. Small molecules facilitate the reprogramming of mouse fibroblasts into pancreatic lineages. *Cell Stem Cell* 2014; 14: 228–36.
- 81 Nie B, Wang H, Laurent T, Ding S. Cellular reprogramming: a small molecule perspective. *Curr Opin Cell Biol* 2012; 24: 784–92.
- 82 Shi Y, Despons C, Do JT, Hahn HS, Schöler HR, *et al*. Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* 2008; 3: 568–74.
- 83 Zhu S, Ambasudhan R, Sun W, Kim HJ, Talantova M, *et al*. Small molecules enable OCT4-mediated direct reprogramming into expandable human neural stem cells. *Cell Res* 2014; 24: 126–9.
- 84 Zhu S, Li W, Zhou H, Wei W, Ambasudhan R, *et al*. Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell* 2010; 7: 651–5.
- 85 Chen J, Liu H, Liu J, Qi J, Wei B, *et al*. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet* 2013; 45: 34–42.
- 86 Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, *et al*. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 2013; 502: 65–70.
- 87 Triezenberg SJ, Kingsbury RC, McKnight SL. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* 1988; 2: 718–29.
- 88 Triezenberg SJ, LaMarco KL, McKnight SL. Evidence of DNA: protein interactions that mediate HSV-1 immediate early gene activation by VP16. *Genes Dev* 1988; 2: 730–42.
- 89 Sadowski I, Ma J, Triezenberg S, Ptashne M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 1988; 335: 563–4.
- 90 Horb ME, Shen CN, Tosh D, Slack JM. Experimental conversion of liver to pancreas. *Curr Biol* 2003; 13: 105–15.
- 91 Kaneto H, Nakatani Y, Miyatsuka T, Matsuoka TA, Matsuhisa M, *et al*. PDX-1/VP16 fusion protein, together with NeuroD or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. *Diabetes* 2005; 54: 1009–22.
- 92 Nagaya M, Katsuta H, Kaneto H, Bonner-Weir S, Weir GC. Adult mouse intrahepatic biliary epithelial cells induced *in vitro* to become insulin-producing cells. *J Endocrinol* 2009; 201: 37–47.
- 93 Wang Y, Chen J, Hu JL, Wei XX, Qin D, *et al*. Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep* 2011; 12: 373–8.
- 94 Hammachi F, Morrison GM, Sharov AA, Livigni A, Narayan S, *et al*. Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. *Cell Rep* 2012; 1: 99–109.
- 95 Aksoy I, Jauch R, Eras V, Chng WB, Chen J, *et al*. Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming. *Stem Cells* 2013; 31: 2632–46.
- 96 Weintraub H. The MyoD family and myogenesis: redundancy, networks, and thresholds. *Cell* 1993; 75: 1241–4.
- 97 Hirai H, Tani T, Katoku-Kikyo N, Kellner S, Karian P, *et al*. Radical acceleration of nuclear reprogramming by chromatin remodeling with the transactivation domain of MyoD. *Stem Cells* 2011; 29: 1349–61.
- 98 Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 1994; 77: 451–9.
- 99 Hirai H, Tani T, Kikyo N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int J Dev Biol* 2010; 54: 1589–96.
- 100 Hirai H, Katoku-Kikyo N, Karian P, Firpo M, Kikyo N. Efficient iPS cell production with the MyoD transactivation domain in serum-free culture. *PLoS One* 2012; 7: e34149.
- 101 Warren L, Ni Y, Wang J, Guo X. Feeder-free derivation of human induced pluripotent stem cells with messenger RNA. *Sci Rep* 2012; 2: 657.
- 102 Hirai H, Katoku-Kikyo N, Keirstead SA, Kikyo N. Accelerated direct reprogramming of fibroblasts into cardiomyocyte-like cells with the MyoD transactivation domain. *Cardiovasc Res* 2013; 100: 105–13.
- 103 Zhu G, Li Y, Zhu F, Wang T, Jin W, *et al*. Coordination of engineered factors with TET1/2 promotes early-stage epigenetic modification during somatic cell reprogramming. *Stem Cell Reports* 2014; 2: 253–61.
- 104 Qing Y, Yin F, Wang W, Zheng Y, Guo P, *et al*. The Hippo effector Yorkie activates transcription by interacting with a histone methyltransferase complex through Nco6. *Elife* 2014; 3.
- 105 Lian I, Kim J, Okazawa H, Zhao J, Zhao B, *et al*. The role of YAP transcription



- coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev* 2010; 24: 1106–18.
- 106 Overholtzer M, Zhang J, Smolen GA, Muir B, Li W, *et al*. Transforming properties of YAP, a candidate oncogene on the chromosome 11q22 amplicon. *Proc Natl Acad Sci U S A* 2006; 103: 12405–10.
- 107 Yagi R, Chen LF, Shigesada K, Murakami Y, Ito Y. A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. *EMBO J* 1999; 18: 2551–62.
- 108 Hauri S, Wepf A, van Drogen A, Varjosalo M, Tapon N, *et al*. Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. *Mol Syst Biol* 2013; 9: 713.
- 109 Marcus GA, Horiuchi J, Silverman N, Guarente L. ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. *Mol Cell Biol* 1996; 16: 3197–205.
- 110 Gaj T, Gersbach CA, Barbas CF 3<sup>rd</sup>. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 2013; 31: 397–405.
- 111 Kim H, Kim JS. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 2014; 15: 321–34.
- 112 Suzuki K, Yu C, Qu J, Li M, Yao X, *et al*. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* 2014; 15: 31–6.
- 113 Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, *et al*. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 2014; 15: 27–30.
- 114 Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, *et al*. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* 2011; 6: e19509.
- 115 Morbitzer R, Römer P, Boch J, Lahaye T. Regulation of selected genome loci using *de novo*-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A* 2010; 107: 21617–22.
- 116 Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, *et al*. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009; 326: 1509–12.
- 117 Bartsevich VV, Miller JC, Case CC, Pabo CO. Engineered zinc finger proteins for controlling stem cell fate. *Stem Cells* 2003; 21: 632–7.
- 118 Juárez-Moreno K, Ericas R, Beltran AS, Stolzenburg S, Cuello-Fredes M, *et al*. Breaking through an epigenetic wall: re-activation of Oct4 by KRAB-containing designer zinc finger transcription factors. *Epigenetics* 2013; 8: 164–76.
- 119 Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, *et al*. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 2012; 40: 5368–77.
- 120 Gao X, Yang J, Tsang JC, Ooi J, Wu D, *et al*. Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. *Stem Cell Reports* 2013; 1: 183–97.
- 121 Weintraub H, Dwarki VJ, Verma I, Davis R, Hollenberg S, *et al*. Muscle-specific transcriptional activation by MyoD. *Genes Dev* 1991; 5: 1377–86.
- 122 Davis RL, Weintraub H. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* 1992; 256: 1027–30.
- 123 De Masi F, Grove CA, Vedenko A, Alibés A, Gisselbrecht SS, *et al*. Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res* 2011; 39: 4553–63.
- 124 Huang J, Weintraub H, Kedes L. Intramolecular regulation of MyoD activation domain conformation and function. *Mol Cell Biol* 1998; 18: 5478–84.
- 125 Meijnsing SH, Pufall MA, So AY, Bates DL, Chen L, *et al*. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 2009; 324: 407–10.
- 126 van de Wetering M, Oosterwegel M, van Norren K, Clevers H. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J* 1993; 12: 3847–54.
- 127 Harley VR, Jackson DI, Hextall PJ, Hawkins JR, Berkovitz GD, *et al*. DNA binding activity of recombinant SRY from normal males and XY females. *Science* 1992; 255: 453–6.
- 128 Jauch R, Ng CK, Narasimhan K, Kolatkar PR. The crystal structure of the Sox4 HM domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. *Biochem J* 2012; 443: 39–47.
- 129 Palasingam P, Jauch R, Ng CK, Kolatkar PR. The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J Mol Biol* 2009; 388: 619–30.
- 130 Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, *et al*. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* 2003; 17: 2048–59.
- 131 Werner MH, Huth JR, Gronenborn AM, Clore GM. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 1995; 81: 705–14.
- 132 Williams DC Jr, Cai M, Clore GM. Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. *J Biol Chem* 2004; 279: 1449–57.
- 133 Ng CK, Li NX, Chee S, Prabhakar S, Kolatkar PR, *et al*. Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Res* 2012; 40: 4933–41.
- 134 Jauch R, Kolatkar PR. What makes a pluripotency reprogramming factor? *Curr Mol Med* 2013; 13: 806–14.
- 135 King CY, Weiss MA. The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proc Natl Acad Sci U S A* 1993; 90: 11990–4.
- 136 Phillips NB, Racca J, Chen YS, Singh R, Jancso-Radek A, *et al*. Mammalian testis-determining factor SRY and the enigma of inherited human sex reversal: frustrated induced fit in a bent protein-DNA complex. *J Biol Chem* 2011; 286: 36787–807.
- 137 Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. *Development* 2013; 140: 4129–44.
- 138 Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet* 2000; 16: 182–7.
- 139 Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* 2013; 12: 15–30.
- 140 Wilson M, Koopman P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev* 2002; 12: 441–6.
- 141 Bernard P, Harley VR. Acquisition of SOX transcription factor specificity through protein-protein interaction, modulation of Wnt signalling and post-translational modification. *Int J Biochem Cell Biol* 2010; 42: 400–10.
- 142 Jauch R, Aksoy I, Hutchins AP, Ng CK, Tian XF, *et al*. Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells* 2011; 29: 940–51.
- 143 Niakan KK, Ji H, Maehr R, Vokes SA, Rodolfa KT, *et al*. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* 2010; 24: 312–26.
- 144 Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, *et al*. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* 2013; 32: 938–53.
- 145 Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, *et al*. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol* 2013; 15: 295–301.
- 146 Merino F, Ng CK, Veerapandian V, Schöler HR, Jauch R, *et al*. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. *Structure* 2014; 22: 1274–86.
- 147 Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, *et al*. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 2012; 150: 1209–22.
- 148 Buganim Y, Faddah DA, Jaenisch R. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet* 2013; 14: 427–39.
- 149 Gonçalves MA, Janssen JM, Nguyen QG, Athanasopoulos T, Hauschka SD, *et al*. Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol Ther* 2011; 19: 1331–41.