Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Synthetic transcription factors for cell fate reprogramming

Joshua B Black[1,2] and Charles A Gersbach[1,2,3]

The ability to reprogram cell lineage specification through the activity of master regulatory transcription factors has transformed disease modeling, drug screening, and cell therapy for regenerative medicine. Recent advances in the engineering of synthetic transcription factors to modulate endogenous gene expression networks and chromatin states have generated a new set of tools with unique advantages to study and enhance cell reprogramming methods. Several studies have applied synthetic transcription factors in various cell reprogramming paradigms in human and murine cells. Moreover, the adaption of CRISPR-based transcription factors for high-throughput screening will enable the systematic identification of optimal factors and gene network perturbations to improve current reprogramming protocols and enable conversion to more diverse, highly specified, and mature cell types. The rapid development of next-generation technologies with more robust and versatile functionality will continue to expand the application of synthetic transcription factors for cell reprogramming.

Addresses
[1] Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA
[2] Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA
[3] Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC 27710, USA

Corresponding author: Gersbach, Charles A (charles.gersbach@duke.edu)

Current Opinion in Genetics & Development 2018, **52**:13–21

This review comes from a themed issue on **Cell reprogramming, regeneration and repair**

Edited by **Knut Woltjen** and **Alex Bortvin**

https://doi.org/10.1016/j.gde.2018.05.001

Introduction

Transcription factors (TFs) transform diverse intracellular and extracellular signals into dynamic patterns of gene expression that define cellular phenotype. During development, they coordinate to initiate and maintain the transcriptional programs specifying the diversity of cell types in an organism. As a result of the natural capacity of TFs to program cell states, ectopic overexpression of TFs has become an effective strategy to reprogram cell phenotypes for desired biomedical applications [1]. A primary example of this is the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) through the overexpression of transgenes encoding Oct4, Sox2, Klf4, and c-Myc [2]. These iPSCs have the capacity for indefinite self-renewal and can be differentiated into cell lineages derived from all three germ layers, establishing a powerful model system for studying disease mechanisms, characterizing drug activity, and developing cell therapies [3].

In addition to the overexpression or direct delivery of naturally occurring TFs, several studies have used synthetic transcription factors (Syn-TFs) to regulate endogenous gene expression in a precise and programmable manner [4]. Syn-TFs consist of a modular DNA-binding domain coupled with a catalytic or scaffold domain that can modulate gene expression when localized to genomic regulatory elements. Certain Syn-TF fusions are designed to deposit or remove precise epigenetic marks at target loci or alter chromatin accessibility and are amenable to high-throughput screens [4–6]. Ultimately, the versatility of Syn-TFs make them uniquely suited for the control of endogenous gene networks and epigenetic landscapes to study and reprogram the role gene expression plays in defining cellular identity. Here, we review the use of Syn-TFs for applications in cellular reprogramming.

Synthetic transcription factors

The three most common DNA-binding platforms used to construct Syn-TFs are zinc finger proteins (ZFPs), transcription activator-like effectors (TALEs), and the clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) system [7]. Although the CRISPR/Cas9 system has recently become the most widely adopted, all three technologies have advantages and limitations making them preferentially suited for particular applications. CRISPR/Cas9 follows a simple design for targeting new sites through Watson-Crick base pairing encoded on a short guide RNA (gRNA) molecule that recruits the Cas9 protein to a genomic target site [8]. Building a new ZFP or TALE is relatively challenging in comparison as it requires further protein engineering to specify DNA binding preference. The CRISPR/Cas9 system is conducive to multiplex targeting of distinct loci through co-delivery of several gRNA cassettes with a single Cas9-based effector. However, a single Cas9-based

effector does not discriminate between co-expressed gRNAs, and thus cannot naturally achieve orthogonal activity. In contrast, ZFPs and TALEs are inherently specific for only their single target site and thus also orthogonal. Multiplexed, orthogonal activities can be realized with CRISPR through the use of Cas9 enzymes from different bacterial species, conditional Cas9 fusions, or engineered gRNAs attached to aptamers that mediate distinct gene regulatory functions [9–11].

Syn-TFs are constructed by fusing gene regulatory domains to ZFPs, TALEs, or nuclease deactivated Cas9 (dCas9). The most commonly used domains to mediate gene activation and repression are oligomers of the VP16 peptide and the KRAB domain, respectively [4]. Tetramers of the VP16 peptide, dubbed VP64 [12], function to recruit the preinitiation complex to gene promoters [13]. The KRAB domain is present in many natural human ZFP-based transcriptional repressors and functions by recruiting KAP1, HP-1, and other heterochromatin-forming factors [14]. In addition to these non-enzymatic scaffold domains, numerous other trans-acting or epigenetic modifying domains have been utilized to deposit or remove certain epigenetic marks or remodel chromosomal topology [15–17,18•]. These Syn-TFs can regulate endogenous gene expression and epigenetic marks from gene-proximal and distal regulatory elements [15,16,19].

Syn-TF platforms can be adapted for high-throughput screening of thousands of Syn-TF variants [5,6,20,21]. ZFPs are uniquely suited for screening applications, as target site length can be shortened to increase binding frequency across the genome, enabling genome-wide transcriptional transformations similar to natural transcription factors [20,21,22•]. However, identifying the functional targets of such a ZFP can require laborious experimental and bioinformatic validations. The CRISPR system is amenable to high-throughput screening with pooled libraries of gRNAs synthesized as short oligonucleotides [23,24]. In this case, the integrated gRNA sequence within a cell serves as a barcode for the targeted genomic region. These screens are particularly useful for applications where there is minimal prior knowledge regarding the functional genes and regulatory elements mediating a particular phenotype.

Recently it has been shown that Cas9 activity is sensitive to nucleosome positioning, and algorithms to predict optimal gRNA activity account for nucleosome occupancy to maximize Cas9 efficacy [25–27]. Even so, Syn-TFs can function analogously to pioneer transcription factors in that they can bind regions of heterochromatin, initiate chromatin remodeling, and enhance accessibility for other DNA-binding factors [28,29]. These characteristics enhance the versatility of Syn-TFs and are essential to enable precise manipulation of endogenous transcriptional networks for cell reprogramming applications.

Synthetic transcription factors for cellular reprogramming

Overexpression of naturally occurring transcription factors has proven to be an effective strategy to reprogram somatic cells [2,30,31]. It is generally considered necessary to overexpress these TFs at supraphysiological levels in order to overcome barriers to activity, such as a repressive chromatin environment and the limited availability of cofactors [32]. Consequently, most reprogramming protocols rely on lentiviral delivery of transgenes encoding TFs driven by strong promoters for elevated and prolonged expression. Given that Syn-TFs are unlikely to induce levels of endogenous gene expression comparable to that from ectopic transgenes, it was uncertain if targeted gene activation with Syn-TFs could be sufficiently robust to mediate such a substantial cellular transformation.

Early studies used Syn-TFs to target *Oct4*, a key regulator of stem cell identity, to modulate differentiation in embryonic stem cells or induce pluripotency in differentiated cell types [33–35]. Gao *et al.* targeted TALE-VP64 to an enhancer of *Oct4* concurrent with overexpression of SOX2, KLF4, and C-MYC (SKM) to reprogram mouse embryonic fibroblasts (MEFs) to iPSCs [35]. Targeting the *Oct4* enhancer, as compared to OCT4 overexpression, more rapidly activated endogenous *Oct4* expression and remodeled the chromatin environment surrounding the *Oct4* locus. This led to a higher number of Rex1-positive colonies observed early in reprogramming, although this was eventually surpassed by OCT4 overexpression at later stages. While this method required SKM co-expression, it was an influential proof-of-principle study that established Syn-TFs as robust tools capable of mediating cellular reprogramming.

The differences in chromatin remodeling and reprogramming kinetics between TALE-VP64-mediated *Oct4* activation compared to OCT4 overexpression evoked the hypothesis that directly modifying the endogenous gene might establish a more accessible chromatin environment for co-occupancy of other factors to enhance *Oct4* expression and the pluripotent phenotype. In a subsequent study, the same authors revealed that TALE-based activators outperformed dCas9-based activators in the same iPSC reprogramming model, speculating that dCas9 interfered with binding of endogenous factors and stabilization of expression late in reprogramming [36]. However, a different group later showed that dCas9-based activation of *Oct4* was sufficient to replace ectopic OCT4 in reprogramming human cells to pluripotency [37].

One of the first examples to exclusively use a Syn-TF to achieve reprogramming was in the direct conversion of MEFs to skeletal myocytes through targeted activation of endogenous *Myod1* via a dCas9-based activator with VP64 domains on both the N-terminus and C-terminus (^{VP64}dCas9^{VP64}) [38]. In addition to generating multinucleated myotubes at similar kinetics and efficiency as with