Published in final edited form as:

Biochem J. 2014 September 15; 462(3): 397–413. doi:10.1042/BJ20140400.

Controlling gene networks and cell fate with precision-targeted DNA-binding proteins and small-molecule-based genome readers

Asuka Eguchi*, Garrett O. Lee*, Fang Wan†, Graham S. Erwin‡, and Aseem Z. Ansari†,‡,1

*Graduate Program in Cellular and Molecular Biology, University of Wisconsin-Madison, 1525 Linden Drive, Madison, WI 53706, U.S.A

†The Genome Center of Wisconsin, University of Wisconsin-Madison, 425 Henry Mall, Madison, WI 53706, U.S.A

‡Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, U.S.A

Abstract

Transcription factors control the fate of a cell by regulating the expression of genes and regulatory networks. Recent successes in inducing pluripotency in terminally differentiated cells as well as directing differentiation with natural transcription factors has lent credence to the efforts that aim to direct cell fate with rationally designed transcription factors. Because DNA-binding factors are modular in design, they can be engineered to target specific genomic sequences and perform pre-programmed regulatory functions upon binding. Such precision-tailored factors can serve as molecular tools to reprogramme or differentiate cells in a targeted manner. Using different types of engineered DNA binders, both regulatory transcriptional controls of gene networks, as well as permanent alteration of genomic content, can be implemented to study cell fate decisions. In the present review, we describe the current state of the art in artificial transcription factor design and the exciting prospect of employing artificial DNA-binding factors to manipulate the transcriptional networks as well as epigenetic landscapes that govern cell fate.

Keywords

artificial transcription factor; cell fate; genome editing; reprogramming; synthetic biology

INTRODUCTION

The expression of certain TFs (transcription factors) can redirect the dynamics of a cell to dramatically change its fate and function. When Takahashi and Yamanaka [1] reported the ability to reprogramme differentiated somatic cells into iPS (induced pluripotent stem) cells, they reminded us of the determinative power of TFs: the overexpression of TFs alone (Oct4,

Sox2, Klf4 and c-Myc) was sufficient to profoundly alter the fate of a cell. TF-mediated regulation of cell fate perhaps should not have come as such a surprise because Weintraub and colleagues demonstrated in 1987 that the overexpression of a single TF, MyoD, is sufficient to transdifferentiate fibroblasts to myocytes [2]. Several subsequent studies have identified such key regulatory TFs in specific stages of cellular differentiation [3,4]. Reprogramming differentiated cells to a pluripotent state, however, has inspired the field to focus on identifying key TFs that alter a cell's phenotype. Since the first report of the Yamanaka factors [1], other combinations of factors (Oct4, Sox2, Lin28 and Nanog [5]; Oct4, Sox2 and Esrrb [6]; Oct4, Sox2, Klf4 and Tbx3 [7]) as well as small molecules [8-11] have been used for reprogramming. These reprogramming factors maintain the stem cell phenotype primarily by up-regulating pluripotency and self-renewal genes and down-regulating lineage commitment genes [12].

TFs have also been applied to stem cells to differentiate them to a desired cell type. During the differentiation process, cell fate becomes progressively restricted as pluripotent stem cells give rise to multipotent progenitors, which in turn give rise to differentiated cells [13]. Of the lineage commitment studies, haemopoiesis has been the best characterized [14]. However, progress in directed differentiation has been time-consuming and cost-prohibitive because these studies involve testing a host of candidate factors obtained by studying embryonic development or by comparing expression profiles of different cell types [3]. This trial-and-error approach may, in some instances, produce the desired cell type; however, maintenance factors that sustain a given cell type tightly regulate molecular circuits and the chromatin environment to achieve a homeostatic state. Overcoming this homeostatic state remains an obstacle for manipulating cell fate efficiently and reliably.

Furthermore, several studies have revealed that the transcriptome and epigenetic landscape between ES (embryonic stem) and iPS cells differ with the latter bearing some memory of its original differentiated state [7,15-19]. Investigating the reported epigenetic and transcriptional differences between cells manipulated *in vitro* to those found *in vivo* as well as uncovering the detailed transcriptional networks that regulate cell fate will be important for therapeutic purposes in regenerative medicine, disease modelling and autologous cell-based therapies [20].

ATFs (artificial TFs) and designer enzymes, which function independently of cellular states and signals [21], are emerging as exciting tools to modify cell fate in the context of stem cells. Although inspired by natural TFs, ATFs can be designed to control transcription in ways that natural TFs cannot: the function, timing, concentration and ability to interact with partner proteins and signals can be engineered to provide a high degree of external control. ATFs can be designed to up-regulate or down-regulate specific genes without relying on partner proteins that may be necessary for natural TFs. Small molecule ATFs, which do not leave a genetic footprint, are an appealing class of transcriptional regulators with a promising potential to serve as therapeutics. For the purposes of down-regulating gene expression, ATFs can be preferable to RNAi, which relies on the RNAi machinery, is restricted from certain tissues and is not effective when the protein has a long half-life [22]. Furthermore, chromatin-remodelling enzymes can be attached to engineered DBDs (DNA-binding domains) to regulate epigenetic modifications in a site-specific manner [23-26]. In a

related approach, genome editing can be performed with designer nucleases followed by homologous recombination to insert or delete genes at specific loci.

The present review will first cover the state of the art design principles of DNA-binding proteins and genome-targeting small molecules ('Toolbox and modular design' section). In the 'Applications in controlling gene networks' section, we address the exciting application of these tools through regulatory control or permanent changes to the genome for the purpose of directing cell fate decisions. New ways in which to apply the technologies described in the 'Toolbox and modular design' section will also be covered. By thoughtful application of artificial DNA-binding factors and small molecules, the transcriptional network and epigenetic landscape of cells can be perturbed in a targeted manner to obtain unprecedented insights as well as exquisite control of the regulatory events that govern cell fate.

TOOLBOX AND MODULAR DESIGN

Many eukaryotic proteins, including TFs, are modular in design (Figure 1A). TFs typically comprise a DBD, an ID (interaction domain) and an ED (effector domain), each of which can be customized such that the modularly assembled artificial factor targets a specific genomic sequence and functions in a predetermined way.

The DBD can be designed to target genomic sites of particular length and sequence. The human genome is 3 billion bp in size; therefore a 16 bp target would, theoretically, occur uniquely in the genome. However, most eukaryotic TFs target 8–10 bp sites and can bind thousands of sites across the genome. To target specific genes, TFs interpret information embedded in the genome by binding 'combinatorially' with other TFs that are selectively mobilized by different cellular signals [27,28]. In other words, TFs integrate cellular signals such that specific combinations 'co-operatively' associate with different sites to elicit appropriate transcriptional responses [29–31].

In addition to transient signal-responsive gene regulation, the balance of different TFs can lead to bistable gene switches that stabilize cellular states and lineage-specific transcriptional circuits [32]. Co-operative assembly between TFs is often achieved through IDs. Mimicking natural design, ATFs can be engineered to interact with natural TFs, thereby integrating artificial molecules into cellular circuits as well as enhancing specificity in targeting relevant genomic loci. Finally, the ED should be chosen such that the engineered factor functions as an activator, repressor or enzyme with a desired catalytic function that enables editing of epigenetic marks on histones and DNA or even editing of genomic content (Figure 1B).

DBDs

The DBD plays a major role in targeting genomic loci by making sequence-specific contacts with DNA and even other natural TFs that bind juxtaposed sites. Various binding domains, such as ZFs (zinc fingers) (Box 1), TALEs (transcription activator-like effectors) (Box 2), as well as non-protein based synthetic molecules (Box 3), have been engineered to create customized factors with desired genome-targeting properties. Until fairly recently, ZFs were

widely used for gene regulation and genome editing with nucleases. Customized DNA-binding proteins are now increasingly being constructed with TALEs, which are more amenable to rational design. DNA targeting is even being achieved by RNA-guided CRISPR–Cas proteins (clustered regularly interspaced short palindromic repeats–CRISPR-associated) (Box 4, left-hand side), which are known for their ease of design. Over the years, much effort has gone into designing DNA-binding molecules with desired sequence recognition properties. These efforts can be grouped in two classes: those focused on engineering natural DBDs and those focused on small molecule genome readers.

Targeting specific sites in the genome can be achieved by selecting the appropriate DBD. Each DBD shown in the boxes has its own set of advantages and disadvantages. TALEs are tailored for targeting unique sites in the genome and can be rationally designed with minimal off-target effects. However, TALEs are inherently large proteins, so other DBDs may be more suited for delivery into mammalian cells. In addition to the discussed domains, miniature proteins derived from combinatorial protein evolution of homeodomains can be used to satisfy size constraints [33–35]. Because the nuclease activity of CRISPR–Cas is so efficient, this system may be most useful for permanent genome editing.

Non-protein-based ATFs may be most appropriate when transient gene regulation is desired without use of genetically encoded material. Oligonucleotides, oligonucleotide analogues and polyamides are examples of such synthetic molecules that target specific DNA sites [36–38]. TFOs (triplex-forming oligonucleotides) and their analogues bind DNA by making Hoogsteen hydrogen bonds in the major groove [39]. PNAs (peptide nucleic acids) are another flavour of oligonucleotide analogues in which a synthetic peptide backbone replaces the sugar phosphate backbone of the nucleic acids [40]. Their stability and resistance to enzymatic degradation in cells have made PNAs attractive DNA-binding modules in designing synthetic ATFs. Of synthetic DBDs, minor groove-binding polyamides have emerged as the most reliable class of molecules that can be rationally designed to target desired DNA sequences [41]. Although polyamides can serve as effective DBDs, there still lies a challenge in delivery to the nucleus. However, conjugation to IPA (isophthalic acid) or incorporation of aryl turns, both of which enhance cellular uptake, may resolve this problem [42,43].

The movement towards inducing cell fate conversion entirely with small molecules without leaving a genetic footprint underscores the importance of synthetic ATFs [10,44,45]. Moreover, some regions of the genome may not be accessible to protein-based DBDs. Highly chromatinized and silenced regions of the genome can be targeted with synthetic genome readers such as polyamides, which have the ability to bind to chromatinized and methylated repressed regions while maintaining their rationally designed sequence specificity [46,47].

Co-operative and combinatorial control through engineered IDs

The ID enables DNA-binding factors to co-operatively assemble with other proteins (TFs or enzymes) at desired binding sites (Figure 2A). The incorporation of IDs in artificial factors is an obvious means to improve DNA affinity and specificity. Because ATFs are modular by design, each domain can be derived from proteins or small molecules, as demonstrated by

conjugating the coiled-coil (i.e. leucine zipper) dimerization region of Gcn4 (residues 251–281) to a hairpin polyamide to generate the first synthetic ATF [113,130]. Furthermore, fusions of leucine zipper IDs to ZF DBDs produced homodimers (GCN4 pair) and heterodimers (Fos–Jun) [131,132]. Peptides selected by phage display [133,134], coiled-coiled interactions characterized by protein microarrays [135], as well as small molecules [136], can be attached to ATFs to recruit other natural factors or ATFs. Even short rationally designed dipeptides, such as the ‘WM’ motif used to recruit the Exd (extradenticle) homeodomain, function as IDs (Figures 2B and 2C) [99,114,137].

An added layer of temporal and induction level control can be included by using synthetic linkers that enable interactions at physiological temperatures, but are disruptive at higher temperatures [114]. These synthetic temperature-sensitive derivatives can be further elaborated to yield small molecule or secondary metabolite-responsive synthetic IDs. Other classical approaches include fusion of the LBD (ligand-binding domain) of the oestrogen receptor to gene-activating ZF ATFs that could dimerize through the LBD and activate transcription in a 4-hydroxytamoxifen-dependent manner [138]. For designer nucleases, heterodimerization through the FokI endonuclease domain has been used to avoid off-target effects, and catalytic domains have been engineered to facilitate heterodimerization [139].

The advantage of using an ID is the co-operativity that comes with the assembly of two or more factors to DNA to regulate gene expression. When two ATFs, each of which target a short sequence, dimerize, the complex as a whole targets a larger binding site, and thus has greater affinity and specificity for targeted DNA sites. The longer dwell time on DNA will also contribute to increased induction of the target gene (if the ATF is designed with an activation domain) because there is more time to engage the general transcriptional machinery. Furthermore, ATFs can be designed to interact with endogenous proteins expressed in the cell such that the binding partners can contribute to DNA-targeting specificity. Moreover, if the endogenous partners are expressed in specific cell types or mobilized in response to specific signals, the interacting ATFs would function in a cell-type-specific and/or signal-responsive manner. Thus thoughtful design and incorporation of IDs can provide ATFs that integrate into the natural regulatory circuitry and provide nuanced control of target genes and their networks.

EDs

The ED determines the output function of the protein upon binding to DNA. An activation domain can be used to recruit the general transcriptional machinery (Figure 3). Conversely, a repression domain can bring in machinery to occlude TF binding sites or recruit complexes that make the local chromatin structures more repressive (Figure 4). Alternatively, an enzymatic domain can be used to make epigenetic modifications at desired sites for an additional layer of control. These catalytic domains can be directed towards making post-translational modifications on histone tails or altering the methylation profile of cytosines in the genome. Furthermore, a nuclease domain can be used as the ED to create a sequence-specific break in DNA for genome-engineering purposes (Box 4). Editing the genome can be a powerful way to create genetic knockins or knockouts important for deciphering how specific genes affect cell fate decisions. Also SNPs (single nucleotide polymorphisms)

associated with diseased states can be corrected or created by genome engineering to study the effects of such mutations on the aetiology of various spontaneous or inherited genetic diseases.

The activation and repression domains described in detail are typically used in the modular assembly of ATFs; however, the same domains can be fused to natural TFs to make them more potent. For example, the transactivation domains of MyoD (residues 3–56) and VP16 have each been attached to Oct4 to make it a more robust activator of the pluripotency circuitry, thereby increasing the number of iPS cell colonies produced [140,141]. VP64 and KRAB (Krüppel-associated box) are the most commonly used activation and repression domains respectively; however, the degree to which these domains can induce or silence transcription depends on the target, and in some instances, another activation or repression domain could be more effective due to the local chromatin structure and co-operative binding of other protein complexes at that site [142,143].

APPLICATIONS IN CONTROLLING GENE NETWORKS

Artificial DNA-binding factors can be used in a number of ways to dissect transcriptional networks regulating cell fate decisions or to create cell lines that are more conducive for a particular scientific or clinical purpose. Below, we discuss how ATFs have been used as effective tools to regulate specific genes [119,158,163-168] or to induce specific cellular phenotypes [112,141,169-175]. In the ‘Precision-targeted epigenetic modifiers’ section, we describe epigenetic modifiers, which can be used to remodel the chromatin structure at specific genomic locations [176-179]. In the ‘Gene regulation through genome editing: sequence-targeted nucleases’ section, we cover gene regulation through genome editing. ZFNs (ZF nucleases), TALENs (TALE nucleases) and CRISPR–Cas have been used to knock in genetic elements [180], knock out specific genes [168], create specific mutations in isogenic lines [181-183], correct mutations in diseased cells [184] and create invaluable ‘reporter’ cell lines for lineage-specific genes [180,183,185-188]. The potential uses of engineered DNA binders in the context of stem cells have yet to be explored to their full extent. In this section, we discuss strategies that are being implemented and future novel applications that would push the frontiers of stem cell research.

ATFs

ATFs, encoded to function as activators or repressors, are versatile tools to study modulation of gene expression in cells [31,53,115,189]. ATFs function independently of cellular signals and are not restricted by the endogenous circuitry [21]. Because ATFs are modular by design, they can be engineered to be more potent than their natural counterparts [140,141]. Although overexpression of a natural TF may result in redirecting binding partners to alternate binding sites [190,191], ATFs that can function independently of limiting cofactors do not have such sequestering effects to indirectly affect the expression of unintended targets. Underscoring the importance of multi-TF assemblage at regulatory elements, the expression of multiple ATFs targeted to clustered gene regulatory elements results in synergistic effects [167,192].

Taking into consideration the chromatin state and epigenetic barriers, ATFs designed to target DNase hypersensitive or nucleosome-free regions lead to more successful binding events in the genome [118,145,167,175,193,194]. To facilitate ATF access to genomic DNA, a NLS (nuclear localization signal) such as the PKKKRKV sequence from the simian virus 40 large T-antigen can be included [195]. Although most ATFs are designed to function in the nucleus, mitochondrial DNA can be targeted by adding a nuclear export signal and mitochondrial targeting sequence [179,196]. Targeting mitochondrial DNA offers an interesting opportunity to manipulate metabolic pathways that are important in cell fate conversions [197,198].

There are at least two ways in which genes can be regulated by ATFs: a targeted approach and a library-screening approach. The targeted gene regulation involves selection of a specific gene and design of an ATF that specifically binds the relevant promoter or regulatory enhancer elements. The library-screening approach relies on examining many ATFs to identify members that can switch on a lineage-specific gene of interest or induce a certain cellular phenotype. The library-screening approach serves as a versatile genetic screen of the regulatory networks without permanent mutagenesis of specific genes. Each approach has its advantages and disadvantages, which should be considered for desired applications.

Depending on the goals of the experiment, transient, inducible or long-term expression of the ATF may be desirable. For mechanistic studies, retroviral or lentiviral vectors that integrate into the genome may be preferable because they are a more efficient means of delivery for long-term expression, but if patient-derived cells are ultimately going to be used for therapeutic purposes, transgene-free methods such as small-molecule-based synthetic ATFs or genetically encoded ATFs delivered via transposons, *loxP*-flanked vectors, episomal vectors, adenoviruses, minicircles and RNA should be considered [199-201]. For a comparison of delivery methods, see other reviews [11,20].

For reliable expression of encoded ATFs, bidirectional promoters are effective for overcoming epigenetic silencing and maintaining stable expression of transgenes because they mimic regions of active chromatin [202]. Often, the constitutively active EF1 α (elongation factor-1 α) promoter is used as a strong and stable promoter for long-term culture of mammalian cells [203-205]. Inducible systems that incorporate tetracycline-responsive promoters [206] or ligand-dependent domains within the protein can add temporal and dosage control to the experiment [75,138,207]. Even light-sensitive domains can enable temporal control as demonstrated with cryptochrome 2 protein that is sensitive to blue light and can function as an ID to recruit the ED when exposed to blue light [208]. Delivery method, duration of expression and promoter strength should all be considered when designing ATFs.

Targeted gene regulation—Heterochromatin and methylated DNA are major barriers to directly targeting desired regulatory elements with ATFs. When targeting a specific gene, DNase hypersensitivity-based targeting of accessible chromatin is useful for targeting a specific gene [145,194,209] (Figure 5). To facilitate access to chromatin, the Blancafort group combined the methyltransferase inhibitor, 5-azaC (5-aza-2'-deoxycytidine), and the

HDAC (histone deacetylase) inhibitor, SAHA (suberoylanilide hydroxamic acid), with a six-ZF protein targeting the activation of maspin (mammary serine protease inhibitor), a tumour suppressor gene [210]. When the ATF alone was expressed, there was a 70-fold increase in maspin transcripts compared with the untransduced control. The role of general epigenetic perturbation by 5-azaC and SAHA alone on maspin activation was minimal. In contrast, the addition of the inhibitors in concert with the ATF increased transcripts 600-fold, suggesting that there was a synergistic effect between the ATF and the inhibitors. Whether the co-treatment with nonspecific chromatin-targeting small molecules is entirely necessary remains to be determined.

The discovery that DNA-binding factors are necessary and sufficient to generate low-methylated regions, suggests that the local accumulation of ATFs at a particular locus could result in gene activation even for genes with methylated promoters [211]. Alleles silenced by methylation such as the paternal allele for H19 and the maternal allele for IGF2 (insulin-like growth factor 2) can be activated or repressed with ZF ATFs, demonstrating the potential for ATFs to overcome repressive epigenetic marks and serve as therapeutic agents in diseases caused by such repression [212].

When perturbing transcriptional networks, the standard approach of overexpressing a protein of interest can have confounding outcomes. Natural TFs typically function in more than one pathway, and overexpression could affect multiple networks unintentionally. The TF could also disrupt the stoichiometric balance of other proteins, compete for a low abundance shared subunit that is used by multiple different complexes, or sequester a binding partner from functioning in other biological contexts [191]. By contrast, ATFs can provide external control by activating or repressing a specific node of a transcriptional network without any cross-talk with other circuits [213].

For diseases in which the expression of a protein is mis-regulated, delivering an ATF that alters transcription of the corresponding gene to normal levels may be beneficial for therapeutic purposes. For example, β -thalassaemia and sickle cell anaemia can be treated by up-regulating γ -globin [214]. An ATF that increases γ -globin transcripts may be a more effective strategy to treating the diseases than the overexpression of a natural TF that up-regulates the same gene [194].

Instead of overexpression, a major advantage of regulation through natural promoters or enhancers of the targeted protein is that spliceforms can be generated at native stoichiometry. The biological consequence of spliceform balance was highlighted in a mouse ear angiogenesis model [215]. Overexpressing the splice variant *Vegfa*₁₆₄ resulted in angiogenesis accompanied with spontaneous haemorrhages, but an ATF that targets endogenous *Vegfa* induced angiogenesis without haemorrhages because the other splice variants, *Vegfa*₁₂₀ and *Vegfa*₁₈₈, were also produced.

In the context of reprogramming to iPS cells, master regulators of pluripotency have been targeted with ATFs. ZF ATFs targeting Oct4 and Sox2 and a TALE ATF targeting the *miR-302/367* cluster have been fused to a KRAB repression domain to demonstrate that the repression of these targets results in loss of proliferative capacity or pluripotency phenotype

[168,169,216]. Conversely, ZFs and TALEs, targeting the promoters of Oct4, Sox2, Klf4 and c-Myc (referred to as OSKM), have been fused to activation domains (either VP16 or VP64) to increase the expression levels of these pluripotency factors [166,169,216]. A TALE–VP64 fusion targeting the enhancer for Oct4 was used to replace Oct4 in the pluripotency cocktail to reprogramme fibroblasts to iPS cells at ~60 % efficiency compared with the OSKM control [175].

Homeobox proteins such as Oct4 can transform cells to a cancerous phenotype when overexpressed [217]; therefore an inducible ATF that targets Oct4 may provide an approach towards reducing the formation of non-iPS cell colonies that proliferate, but do not exhibit the potential to differentiate into the three germ layers. Titrating an inducible ATF targeting Oct4 over different concentrations such that the optimal level of Oct4 is achieved can be a strategy used towards this end. In an alternative strategy for reprogramming, an activation domain was fused to the pluripotency factors to make them more potent for increasing the efficiency of reprogramming [141]. When tested in mouse embryonic fibroblasts, Oct4–VP16, Sox2–VP16 and Nanog–VP16 fusion proteins performed better than their natural counterparts by increasing the efficiency of iPS cell colony formation 100-fold.

In many ways, ATFs designed to target specific genes are more favourable to expressing natural TFs that have the same targets. ATFs can be designed to activate or repress key nodes of a transcriptional network without perturbing multiple genes inadvertently. While ATFs can be designed to interact with endogenous proteins in the cell, they can also be designed to function independently. With ATFs, splice variants of the target gene can be expressed at natural stoichiometries, which can be biologically meaningful in some contexts. When siRNAs do not work or have non-specific effects, ATFs fused to a repression domain may be a viable option. By controlling the ATF under an inducible system, the timing and level of induction or repression of target genes can also be fine-tuned.

Library-screening approaches—Just as genetic screens can be a powerful way to find novel genes associated with a particular phenotype, ATF libraries can serve as a forward genetics approach to uncovering pathways that are related to an interesting biological process. An ATF library can be more versatile than a classical genetic screen because essential genes missed from typical random mutagenesis screens can be captured with an ATF library screen if the level of activation or repression is modulated. An ATF library can be designed to have thousands to millions of different members, each of which targets a different DNA sequence. To date, the most commonly used DBD for ATF libraries is the ZF motif. More recently, ATF libraries were created with TALEs that target 11 bp sequences [218]. The number of possible genomic targets and the variation among targets depends on how the library is designed. Most libraries have been constructed by linking together three to six single ZF modules that are known to bind DNA with an ED (Figure 6). The library is cloned into retroviruses or lentiviruses, which are transduced in a relevant cell type. By screening for a desired phenotype or expression of a lineage-specific gene, researchers can identify cells that have an active ATF. These active ATFs can be validated in cells and for further analysis of their mechanism of action, target genes of these ATFs can be identified by genome-wide methods, namely, ChIP-seq (chromatin immunoprecipitation sequencing) and transcriptional profiling (RNA-seq) [219,220].

A number of phenotypes and genes have been targeted with ATF libraries (Table 1). ATFs that induce drug resistance [171,172,174,221], cell differentiation [170] and antibody production [173] have been identified. Cell-surface markers [68], ICAM-1 (intercellular adhesion molecule-1) [163], VEGF (vascular endothelial growth factor) [60] and SEAP (secreted embryonic alkaline phosphatase) [222] have also served as gene targets. The Barbas group implemented an elegant selection strategy to retrieve functional ATFs that induce γ -globin expression. The viral packaging protein, Gag-Pol, was placed under the γ -globin promoter such that only cells expressing ATFs that drive γ -globin production would subsequently generate viral particles [165].

By swapping the ED for a particular function, a number of variant libraries can be created. For directed differentiation studies, ATF libraries can be an effective way to map out transcriptional networks and to find genes that are associated with cell fate. Due to the nature of ZF motifs that have been characterized thus far, targetable sequences tend to be purine-rich. New ZF, TALE, CRISPR-Cas and DNA-targeting small molecule libraries, such as a polyamide library, that expand the genome sequence targeting repertoire will provide greater regulatory access of the genome and will yield new insights into circuitries that control cellular fate and function.

Precision-targeted epigenetic modifiers

The global epigenetic profiles of pluripotent stem cells are distinguished from those of differentiated cells in that ES and iPS cells have open euchromatin that is more accessible to TFs, whereas lineage commitment is characterized by densely packed heterochromatin that is less accessible except at cell fate-defining loci [223]. During cell fate conversions, TFs are occluded from their binding sites by histones and repressor complexes [224]. The inaccessibility of regulatory elements to TFs can be overcome by sequence-specific chromatin remodelling enzymes capable of reorganizing the epigenetic landscape at the targeted site to a more permissive state. Enzymes that could be fused to the DBD include epigenetic remodellers that function by modifying DNA or by adding post-translational modifications to histones, such as histone acetyltransferases, HDACs, histone methyltransferases and ubiquitin-conjugating enzymes. Such catalytic domains can be attached to the DBD with flexible or rigid peptide linkers as needed [27,225,226].

To reset chromatin and epigenetic states, cells are often treated with HDAC inhibitors, such as VPA (valproic acid), TSA (trichostatin A) and SAHA [11]. These molecules act broadly and, therefore, have pleiotropic effects that result in cellular toxicity. In contrast, a sequence-targeted strategy that relies on engineered DNA binders guiding delivery of the relevant catalytic module or small molecule promises to be more effective in overcoming epigenetic barriers at desired genomic sites.

Sequence-specific polyamides conjugated to SAHA have been shown to activate Oct4 and Nanog [112]. Expression of targeted genes increased approximately 50-fold in mouse embryonic fibroblasts [112], and this increase was concomitant with the appearance of H4Kac (acetylation of lysine of histone H3), H3K9ac, H3K14ac and H3K4me3 (trimethylation of Lys4 of histone H3), marks associated with open chromatin [227]. Sequence-specific small molecule inhibitors and enzymes can have potent effects at lower

concentrations than typically used without a DBD because conjugation to a DBD can lead to increased local concentrations of the HDAC inhibitor at desired loci rather than broad distribution across the genome.

Along the same lines, DNA methylation can be targeted for demethylation and subsequent transcriptional activation. The DNA methyltransferase inhibitor, 5-azaC, is often used to demethylate regulatory elements; however, its effects are due to dilution of methylated DNA over several rounds of cell division, and the drug is highly toxic to mammalian cells [228]. A tailored approach would be to deliver a fusion protein, consisting of an enzyme that has been implicated in active DNA demethylation [229,230] and a DBD that could be rationally designed to direct the demethylase to specific sites in the genome. For example, TDG (thymine DNA glycosylase) fused to the DBD of RelA induced modest effects on DNA demethylation and increased transcription at a RelA target site [24]. Similarly, TALEs fused to the catalytic domain of TET1, which oxidizes 5-methylcytosine to 5-hydroxymethylcytosine, has been used to induce site-specific CpG demethylation on KLF4, RHOXF2 (RhoX homeobox family member 2) and HBB (haemoglobin β) [25].

Conversely, genes that need to be silenced can be down-regulated epigenetically by delivering chromatin remodelling agents that condense chromatin. To this end, the histone demethylase, LSD1, was fused to 40 different TALEs to demonstrate that H3K4me₂, a marker of active enhancers, could be removed at the target sites for those TALEs [26]. Four of these TALE–LSD1 fusions resulted in modest decreases (1.5–3-fold) of transcript levels of nearby genes. Likewise, a four-ZF protein fused to the catalytic domain of DNMT3a (DNA methyltransferase 3a) was used to selectively methylate a CpG apparent in diseased cells due to a SNP in mitochondrial DNA [179]. Expression of this targeted methyltransferase induced 6-fold greater methylation in cells with the SNP compared with cells with wild-type mitochondrial DNA.

Perhaps the combination of precision-targeted epigenetic modifiers and TFs could have synergistic effects such that the chromatin environment mirrors the active or repressive functions of the TF. Targeted epigenetic modifiers could be used to prime the local chromatin environment, and ATFs could be delivered for direct transcriptional control. The idea that the location of super-enhancers, marked by H3K27ac, are cell-type-specific makes epigenetic remodelling in the context of directed differentiation an appealing endeavour [231]. Studies on reorganizing the chromatin landscape would shed light on the biological relevance of epigenetic regulation to prime the cell for transcriptional regulators.

Gene regulation through genome editing: sequence-targeted nucleases

Technology to edit the genome can be added to the toolbox of transcriptional regulation to provide yet another layer of experimental control. For genome editing purposes, the endonuclease domain of FokI is a widely used type II restriction enzyme that cleaves DNA in a sequence-independent manner as a dimer (Box 4, right-hand side) [232]. Because dimerization is required for catalytic activity, two DNA-binding proteins, each targeting flanking sequences of the cleavage site, could be fused with the FokI endonuclease domain to induce a double-stranded break in a specified site [139]. Subsequently, the break is repaired by NHEJ (non-homologous end-joining) or HDR (homology-directed repair) (Box

4, bottom panel) [233]. Of the two repair mechanisms, NHEJ occurs more frequently than HDR because HDR only occurs during late S- and G₂-phases [234].

Meganucleases, ZFNs, TALENs and CRISPR–Cas9 have opened the doors to editing the genome in myriad ways [116,139,235–240]. To achieve success in editing the genome without off-target effects, a pair of co-operatively binding DBDs should provide enough specificity to target unique sites, and further specificity of cleavage can be ensured by dependence on accurate heterodimerization. Techniques for making obligate heterodimers of FokI have been developed to prevent off-target effects of homodimers [139]. When choosing a target, not only should the sequence length be considered, the target sequence should have as little homology as possible to other sequences in the genome [241]. Upon dimerization, one monomer with high DNA-binding affinity can stabilize its partner with low affinity to a binding site containing mismatches [242]. Thus, if available, two moderate affinity monomers that display a high degree of sequence-specificity should be chosen over strong binders that might still bind to related sites. Since off-target effects are dose-dependent, the expression levels of the nucleases should be modulated to minimal levels necessary to induce cleavage [243]. Studies with ZFs as the DBD have shown that using short 4-amino-acid linkers between the DBD and the FokI domain minimizes the promiscuity of the nuclease [225,244].

In a comparison of different DBDs, Mussolino et al. [245] demonstrated that a ZFN designed to target *CCR5* (CCR is CC chemokine receptor) also cleaved 11 % of *CCR2* alleles where there are two mismatches in the binding site, whereas a *CCR5*-specific TALEN cleaved only 1 % of *CCR2* alleles at a site with an equal number of mismatches. TALENs may be much less susceptible to off-target effects; however, thorough studies that measure cleavage at all possible imperfect sites must still be performed to determine the cytotoxic effects.

Off-target effects are higher for CRISPR–Cas9 because specificity is largely determined by eight to 12 nucleotides in the sgRNA (single-guide RNA) [126,127], but bioinformatic tools can be used to find targetable sites [127]. Guide RNAs with a very high or very low GC content do not fare well, and purines in the four nucleotides at the 3' end of the spacer sequence, which base-pairs adjacent to the PAM (protospacer adjacent motif) sequence, can improve affinity to Cas9 [246]. Additionally, specificity can be improved by using Cas9 proteins that function as nickases guided by two unique sgRNA that target the sense and antisense strands of genomic DNA [129,247]. In this strategy, base excision repair corrects for an off-target nick in dsDNA, but two closely offset nicks that would result in a double-stranded break are unlikely to randomly co-occur anywhere in the genome except at the chosen locus. To generate transgenic mice more efficiently, Cas9-encoding mRNA can be injected into a zygote along with sgRNA and donor DNA for targeted homologous recombination [180,183].

The beauty of site-specific genome engineering by HDR is that a mutation can be inserted or corrected at the natural genomic site [139,237]. One way to study transcriptional networks that regulate cell fate is to knock out an important node. Removal of the *miR-302/367* cluster from the genome of human fibroblasts using TALENs resulted in the inability to

generate iPS cell colonies [168]. The reprogramming capacity of knockout cells could be partially rescued through exogenous expression of *miR-302/367*, demonstrating that this miRNA cluster is necessary for reprogramming into pluripotency.

Also, endogenous loci can be targeted to create reporter cell lines that express fluorescent proteins or append epitope tags to desired proteins [185,186,248,249]. Harnessing the ability to knockin reporter genes at specific loci, Goldberg et al. [248] tagged the histone variant H3.3 with epitope using ZFNs and characterized the localization of this histone variant in ES cells and neural progenitor cells. Interestingly, H3.3 placement changes upon differentiation, and enrichment of H3.3 seen at pluripotency genes in ES cells are not observed in neural progenitor cells.

When inserting a transgene, the *AAVS1* locus was demonstrated to be an ideal locus because repressive chromatin marks do not interfere with transgene expression at this locus, and the expression of flanking genes are not enhanced by the insertion of even strong promoters in four human cell types tested [250].

Although sequence-targeted nucleases have been applied in a number of contexts, we highlight examples of cell fate-engineering. The precision with which we can perform mechanistic studies of gene regulatory elements in cells or animals with isogenic backgrounds opens up an exciting realm of possibilities. Specific binding sites for TFs can be mutated such that they no longer bind to the particular promoter or enhancer element. In this context, the TF's role in regulating a particular gene can be characterized without changing the expression levels of that TF. Regulatory elements can be moved from one genomic locus to another to determine whether sequences within a given genomic region are sufficient to induce changes to chromatin structure. Even SNPs implicated in disease [231] can be inserted into the genome to determine whether they are sufficient to create TF binding hot spots or dead zones. Copy number variations can also be tested in addition to the genomic location of the gene. In addition, the expression of lineage-specific genes can be investigated by knocking in a fluorescent reporter at the genomic site of interest. Genome engineering can allow us to design reductionist experiments within the complex setting of a cell to further our understanding of the transcriptional networks to govern cell fate.

FUTURE DIRECTIONS AND CONCLUDING REMARKS

Artificial DNA-binding molecules are powerful tools that will aid in the exploration of cell fate regulation, chromatin accessibility and transcription network dynamics. Sequence-targeted nucleases have become a promising technology for manipulating genes and gene regulatory elements in an otherwise isogenic background. As we learn more about the off-target effects of ZFs, TALEs, CRISPR-Cas9 and polyamides, it may become useful to design multiple factors, which can heterodimerize and capitalize on the strengths of each type of DBD. More insight on mode of recognition for DNA modifications such as 5-methylcytosine and 5-hydroxymethylcytosine would allow us to understand how epigenetic marks are placed and maintained and why these marks can have such a profound impact on development and cellular function. From a clinical perspective, genome-targeting small

molecules hold promise in regulating transcriptional outcomes without leaving a genetic footprint.

In the context of deciphering and guiding cell fate decisions, ATF libraries can be an invaluable tool in mapping out mechanisms underlying cell reprogramming and differentiation. A comparison of pluripotent cell-derived neural progenitor cells, hepatocytes and fibroblasts, with their somatic counterparts revealed that ES or iPS cell-derived cells are more akin to fetal cells than adult cells [251]. In principle, ATFs can be used to alter epigenetic landscapes in a targeted manner or down-regulate genes associated with earlier stages of development so that the cells adopt a more differentiated or mature homeostatic fate.

Given that differentiating cells into a particular cell type, transplanting them into a patient, and relying on the cells to restore function in a failing organ can be exceedingly challenging [20], a more straightforward approach may be to design a cell that can serve as an agent for morphogen delivery to coerce cells in the native niche to convert into specific cell types. Garg et al. [252] synthesized eight TALEs that were orthogonal to the human genome; that is, these TALEs could only activate or repress transcription of a synthetically engineered circuit and did not bind genomic DNA. Synthetic circuits that function within a cell to perform various functions can be powerful tools for regenerative therapy. The power and promise of precision-targeted DNA-binding molecules to guide cell fate in a precisely programmed manner is still in its infancy, but they are rapidly being deployed to study a broad range of phenomena across the biological and medical sciences.

Acknowledgments

We thank Laura Vanderploeg for help with the artwork. We also thank Tim Kamp, Igor Slukvin, Jamie Thomson, Judith Kimble and James Dutton for their helpful discussions. We appreciate the helpful comments from the reviewer and editor in streamlining the review. We apologize to colleagues whose contributions we were unable to cite due to space constraints.

FUNDING

The Ansari laboratory was supported by the NIH (National Institutes of Health) [grant numbers HL099773 and CA133508], the H.I. Romnes Faculty Award, and the W.M. Keck Foundation. A.E. was supported by a Morgridge Biotechnology Wisconsin Distinguished Fellowship Award. F.W. was supported by a UW-Madison Stem Cell and Regenerative Medicine Center Fellowship. G.S.E. was supported by a Molecular Biosciences Training Grant NIH T32 [grant number GM07215].

References

1. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126:663–676. [PubMed: 16904174]
2. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*. 1987; 51:987–1000. [PubMed: 3690668]
3. Cohen DE, Melton D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet*. 2011; 12:243–252. [PubMed: 21386864]
4. Graf T, Enver T. Forcing cells to change lineages. *Nature*. 2009; 462:587–594. [PubMed: 19956253]
5. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007; 318:1917–1920. [PubMed: 18029452]

6. Feng B, Jiang J, Kraus P, Ng J-H, Heng J-CD, Chan Y-S, Yaw L-P, Zhang W, Loh Y-H, Han J, et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor. *Esrrb Nat Cell Biol.* 2009; 11:197–203.
7. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol.* 2010; 28:848–855. [PubMed: 20644536]
8. Zhu S, Li W, Zhou H, Wei W, Ambasadhan R, Lin T, Kim J, Zhang K, Ding S. Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell.* 2010; 7:651–655. [PubMed: 21112560]
9. Schambach A, Cantz T, Baum C, Cathomen T. Generation and genetic modification of induced pluripotent stem cells. *Expert Opin Biol Ther.* 2010; 10:1089–1103. [PubMed: 20528610]
10. Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, Zhao T, Ye J, Yang W, Liu K, et al. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science.* 2013; 341:651–654. [PubMed: 23868920]
11. Okita K, Yamanaka S. Induced pluripotent stem cells: opportunities and challenges. *Philos Trans R Soc Lond B Biol Sci.* 2011; 366:2198–2207. [PubMed: 21727125]
12. Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell.* 2008; 132:567–582. [PubMed: 18295576]
13. Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* 2005; 19:1129–1155. [PubMed: 15905405]
14. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell.* 2008; 132:631–644. [PubMed: 18295580]
15. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009; 462:315–322. [PubMed: 19829295]
16. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiwu O, Richter L, Zhang J, et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell.* 2009; 5:111–123. [PubMed: 19570518]
17. Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods.* 2011; 8:821–827. [PubMed: 21983960]
18. Gore A, Fung H-L, Agarwal S, Antosiewicz-Bourget J, Loh Y-H, Manos PD, Panopoulos AD, Ruiz S, Yu J, Izpisua Belmonte JC, et al. Somatic coding mutations in human induced pluripotent stem cells. *Nature.* 2011; 471:63–67. [PubMed: 21368825]
19. Ji J, Ng SH, Sharma V, Neculai D, Hussein S, Sam M, Trinh Q, Church GM, McPherson JD, Nagy A, Batada NN. Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. *Stem Cells.* 2011; 30:435–440. [PubMed: 22162363]
20. Stadtfeld M, Hochedlinger K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* 2010; 24:2239–2263. [PubMed: 20952534]
21. Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK, Collins JJ. A synthetic biology framework for programming eukaryotic transcription functions. *Cell.* 2012; 150:647–658. [PubMed: 22863014]
22. Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. *Nat Rev Genet.* 2008; 9:554–566. [PubMed: 18521077]
23. Meister GE, Chandrasegaran S, Ostermeier M. Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. *Nucleic Acids Res.* 2010; 38:1749–1759. [PubMed: 20007601]
24. Gregory DJ, Mikhaylova L, Fedulov AV. Selective DNA demethylation by fusion of TDG with a sequence-specific DNA-binding domain. *Epigenetics.* 2012; 7:344–349. [PubMed: 22419066]
25. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol.* 2013; 31:1–9. [PubMed: 23302909]

26. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol.* 2013; 31:1133–1136. [PubMed: 24013198]
27. Ansari AZ, Mapp AK. Modular design of artificial transcription factors. *Curr Opin Chem Biol.* 2002; 6:765–772. [PubMed: 12470729]
28. Ansari AZ. Chemical crosshairs on the central dogma. *Nat Chem Biol.* 2007; 3:2–7. [PubMed: 17173017]
29. Ptashne, M.; Gann, A. *Genes and Signals*. Cold Spring Harbor Laboratory Press; Cold Spring Harbor: 2002. Some notes on higher eukaryotes.
30. Davidson, EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press; New York: 2006. *cis*-Regulatory modules, and the structure/function basis of regulatory logic; p. 31-86.
31. Peterson-Kaufman KJ, Carlson CD, Rodríguez-Martínez JA, Ansari AZ. Nucleating the assembly of macromolecular complexes. *Chembiochem.* 2010; 11:1955–1962. [PubMed: 20812316]
32. Ptashne M. Principles of a switch. *Nat Chem Biol.* 2011; 7:484–487. [PubMed: 21769089]
33. Montclare JK, Schepartz A. Miniature homeodomains: high specificity without an N-terminal arm. *J Am Chem Soc.* 2003; 125:3416–3417. [PubMed: 12643688]
34. Yang L, Schepartz A. Relationship between folding and function in a sequence-specific miniature DNA-binding protein. *Biochemistry.* 2005; 44:7469–7478. [PubMed: 15895990]
35. Zondlo NJ, Schepartz A. Highly specific DNA recognition by a designed miniature protein. *J Am Chem Soc.* 1999; 121:6938–6939.
36. Dervan PB. Molecular recognition of DNA by small molecules. *Bioorg Med Chem.* 2001; 9:2215–2235. [PubMed: 11553460]
37. Liu B, Han Y, Ferdous A, Corey DR, Kodadek T. Transcription activation by a PNA-peptide chimera in a mammalian cell extract. *Chem Biol.* 2003; 10:909–916. [PubMed: 14583257]
38. Beane RL, Ram R, Gabillet S, Arar K, Monia BP, Corey DR. Inhibiting gene expression with locked nucleic acids (LNAs) that target chromosomal DNA. *Biochemistry.* 2007; 46:7572–7580. [PubMed: 17536839]
39. Simon P, Cannata F, Concordet J-P, Giovannangeli C. Targeting DNA with triplex-forming oligonucleotides to modify gene sequence. *Biochimie.* 2008; 90:1109–1116. [PubMed: 18460344]
40. Ray A, Nordén B. Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB J.* 2000; 14:1041–1060. [PubMed: 10834926]
41. Rodríguez-Martínez JA, Peterson-Kaufman KJ, Ansari AZ. Small-molecule regulators that mimic transcription factors. *Biochim Biophys Acta.* 2010; 1799:768–774. [PubMed: 20804876]
42. Meier JL, Montgomery DC, Dervan PB. Enhancing the cellular uptake of Py-Im polyamides through next-generation aryl turns. *Nucleic Acids Res.* 2012; 40:2345–2356. [PubMed: 22080545]
43. Nickols NG, Jacobs CS, Farkas ME, Dervan PB. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.* 2007; 35:363–370. [PubMed: 17175539]
44. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ, Palecek SP. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci U S A.* 2012; 109:E1848–E1857. [PubMed: 22645348]
45. Zhang Y, Li W, Laurent T, Ding S. Small molecules, big roles: the chemical manipulation of stem cell fate and somatic cell reprogramming. *J Cell Sci.* 2012; 125:5609–5620. [PubMed: 23420199]
46. Minoshima M, Bando T, Sasaki S, Fujimoto J, Sugiyama H. Pyrrole-imidazole hairpin polyamides with high affinity at 5'-CGCG-3' DNA sequence; influence of cytosine methylation on binding. *Nucleic Acids Res.* 2008; 36:2889–2894. [PubMed: 18385159]
47. Suto RK, Edayathumangalam RS, White CL, Melander C, Gottesfeld JM, Dervan PB, Luger K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J Mol Biol.* 2003; 326:371–380. [PubMed: 12559907]
48. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet.* 2009; 10:252–263. [PubMed: 19274049]

49. Wolfe SA, Nekludova L, Pabo CO. DNA recognition by Cys₂His₂ zinc finger proteins. *Annu Rev Biophys Biomol Struct.* 2000; 29:183–212. [PubMed: 10940247]
50. Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys₂His₂ zinc finger proteins. *Annu Rev Biochem.* 2001; 70:313–340. [PubMed: 11395410]
51. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO. Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure.* 1996; 4:1171–1180. [PubMed: 8939742]
52. Segal DJ, Dreier B, Beerli RR, Barbas CF. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A.* 1999; 96:2758–2763. [PubMed: 10077584]
53. Sera T. Zinc-finger-based artificial transcription factors and their applications. *Adv Drug Deliv Rev.* 2009; 61:513–526. [PubMed: 19394375]
54. Choo Y, Klug A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A.* 1994; 91:11168–11172. [PubMed: 7972028]
55. Suzuki M, Yagi N. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc Natl Acad Sci U S A.* 1994; 91:12357–12361. [PubMed: 7809040]
56. Jamieson AC, Wang H, Kim SH. A zinc finger directory for high-affinity DNA recognition. *Proc Natl Acad Sci U S A.* 1996; 93:12834–12839. [PubMed: 8917505]
57. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem.* 2001; 276:29466–29478. [PubMed: 11340073]
58. Liu Q, Xia Z, Zhong X, Case CC. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem.* 2002; 277:3850–3856. [PubMed: 11726671]
59. Benos PV, Lapedes AS, Stormo GD. Is there a code for protein-DNA recognition? Probab(istically)? *Bioessays.* 2002; 24:466–475. [PubMed: 12001270]
60. Bae K-H, Kwon YD, Shin H-C, Hwang M-S, Ryu E-H, Park K-S, Yang H-Y, Lee D-K, Lee Y, Park J, et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol.* 2003; 21:275–280. [PubMed: 12592413]
61. Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, Huber A, Koksche B, Barbas CF. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem.* 2005; 280:35588–35597. [PubMed: 16107335]
62. Maeder ML, Thibodeau-Beganny S, Osia A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, et al. Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell.* 2008; 31:294–301. [PubMed: 18657511]
63. Persikov AV, Rowland EF, Oakes BL, Singh M, Noyes MB. Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. *Nucleic Acids Res.* 2013; 42:1497–1508. [PubMed: 24214968]
64. Persikov AV, Singh M. *Denovo* prediction of DNA-binding specificities for Cys₂His₂ zinc finger proteins. *Nucleic Acids Res.* 2013; 42:97–108. [PubMed: 24097433]
65. Carroll D, Morton JJ, Beumer KJ, Segal DJ. Design, construction and in vitro testing of zinc finger nucleases. *Nat Protoc.* 2006; 1:1329–1341. [PubMed: 17406419]
66. Desjarlais JR, Berg JM. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci U S A.* 1992; 89:7345–7349. [PubMed: 1502144]
67. Isalan M, Klug A, Choo Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol.* 2001; 19:656–660. [PubMed: 11433278]
68. Blancafort P, Magnenat L, Barbas CF. Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol.* 2003; 21:269–274. [PubMed: 12592412]

69. Meng X, Brodsky MH, Wolfe SA. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol.* 2005; 23:988–994. [PubMed: 16041365]
70. Liu J, Stormo GD. Context-dependent DNA recognition code for C₂H₂ zinc-finger transcription factors. *Bioinformatics.* 2008; 24:1850–1857. [PubMed: 18586699]
71. Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods.* 2011; 8:67–69. [PubMed: 21151135]
72. Choo Y, Isalan M. Advances in zinc finger engineering. *Curr Opin Struct Biol.* 2000; 10:411–416. [PubMed: 10981627]
73. Moore M, Klug A, Choo Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A.* 2001; 98:1437–1441. [PubMed: 11171969]
74. Klug A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem.* 2010; 79:213–231. [PubMed: 20192761]
75. Beerli RR, Schopfer U, Dreier B, Barbas CF. Chemically regulated zinc finger transcription factors. *J Biol Chem.* 2000; 275:32617–32627. [PubMed: 10924515]
76. Beerli RR, Barbas CF. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol.* 2002; 20:135–141. [PubMed: 11821858]
77. Boch J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol.* 2010; 48:419–436. [PubMed: 19400638]
78. Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu J-K, Shi Y, Yan N. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science.* 2012; 335:720–723. [PubMed: 22223738]
79. Mak AN-S, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science.* 2012; 335:716–719. [PubMed: 22223736]
80. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-Type III effectors. *Science.* 2009; 326:1509–1512. [PubMed: 19933107]
81. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science.* 2009; 326:1501–1501. [PubMed: 19933106]
82. Cong L, Zhou R, Kuo Y-C, Cunniff M, Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun.* 2012; 3:968. [PubMed: 22828628]
83. Streubel J, Blücher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol.* 2012; 30:593–595. [PubMed: 22781676]
84. Szurek B, Rossier O, Hause G, Bonas U. Type III-dependent translocation of the *Xanthomonas* AvrBs3 protein into the plant cell. *Mol Microbiol.* 2002; 46:13–23. [PubMed: 12366827]
85. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol.* 2010; 29:143–148. [PubMed: 21179091]
86. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science.* 2011; 333:1843–1846. [PubMed: 21960622]
87. Arcamone F, Penco S, Orezzi P, Nicoletta V, Pirelli A. Structure and synthesis of distamycin A. *Nature.* 1964; 203:1064–1065. [PubMed: 14223076]
88. Wartell RM, Larson JE, Wells RD. The compatibility of netropsin and actinomycin binding to natural deoxyribonucleic acid. *J Biol Chem.* 1975; 250:2698–2702. [PubMed: 164447]
89. Finlay AC, Hochstein FA, Sobin BA, Murphy FX. Netropsin, a new antibiotic produced by a *Streptomyces*. *J Am Chem Soc.* 1951; 73:341–343.
90. Wade WS, Mrksich M, Dervan PB. Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J Am Chem Soc.* 1992; 114:8783–8794.

91. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A*. 1985; 82:1376–1380. [PubMed: 2983343]
92. Lown JW, Krowicki K, Bhat UG, Skorobogaty A, Ward B, Dabrowiak JC. Molecular recognition between oligopeptides and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry*. 1986; 25:7408–7416. [PubMed: 3026455]
93. Pelton JG, Wemmer DE. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A*. 1989; 86:5723–5727. [PubMed: 2762292]
94. Trauger JW, Baird EE, Dervan PB. Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA. *Chem Biol*. 1996; 3:369–377. [PubMed: 8807866]
95. Dervan PB, Edelson BS. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol*. 2003; 13:284–299. [PubMed: 12831879]
96. Mrksich M, Dervan PB. Design of a covalent peptide heterodimer for sequence-specific recognition in the minor groove of double-helical DNA. *J Am Chem Soc*. 1994; 116:3663–3664.
97. Chenoweth DM, Dervan PB. Structural basis for cyclic Py-Im polyamide allosteric inhibition of nuclear receptor binding. *J Am Chem Soc*. 2010; 132:14521–14529. [PubMed: 20812704]
98. Bremer RE, Baird EE, Dervan PB. Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Chem Biol*. 1998; 5:119–133. [PubMed: 9545429]
99. Moretti R, Donato LJ, Brezinski ML, Stafford RL, Hoff H, Thorson JS, Dervan PB, Ansari AZ. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein–DNA assembly. *ACS Chem Biol*. 2008; 3:220–229. [PubMed: 18422304]
100. Gottesfeld JM, Melander C, Suto RK, Raviol H, Luger K, Dervan PB. Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J Mol Biol*. 2001; 309:615–629. [PubMed: 11397084]
101. Edayathumangalam RS, Weyermann P, Gottesfeld JM, Dervan PB, Luger K. Molecular recognition of the nucleosomal ‘supergroove’. *Proc Natl Acad Sci U S A*. 2004; 101:6864–6869. [PubMed: 15100411]
102. Lee M, Roldan M, Haskell M, McAdams S, Hartley J. *In vitro* photoinduced cytotoxicity and DNA binding properties of psoralen and coumarin conjugates of netropsin analogues: DNA sequence-directed alkylation and cross-link formation. *J Med Chem*. 1994; 37:1208–1213. [PubMed: 8164263]
103. Best TP, Edelson BS, Nickols NG, Dervan PB. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A*. 2003; 100:12063–12068. [PubMed: 14519850]
104. Arndt H-D, Hauschild KE, Sullivan DP, Lake K, Dervan PB, Ansari AZ. Toward artificial developmental regulators. *J Am Chem Soc*. 2003; 125:13322–13323. [PubMed: 14583004]
105. Olenyuk BZ, Zhang G-J, Klco JM, Nickols NG, Kaelin WG, Dervan PB. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc Natl Acad Sci U S A*. 2004; 101:16768–16773. [PubMed: 15556999]
106. Chou CJ, O’Hare T, Lefebvre S, Alvarez D, Tyner JW, Eide CA, Druker BJ, Gottesfeld JM. Growth arrest of BCR-ABL positive cells with a sequence-specific polyamide-chlorambucil conjugate. *PLoS ONE*. 2008; 3:e3593. [PubMed: 18974832]
107. Xiao X, Yu P, Lim H-S, Sikder D, Kodadek T. A cell-permeable synthetic transcription factor mimic. *Angew Chem Int Ed Engl*. 2007; 46:2865–2868. [PubMed: 17351994]
108. Edwards TG, Koeller KJ, Slomczynska U, Fok K, Helmus M, Bashkin JK, Fisher C. HPV episome levels are potently decreased by pyrrole-imidazole polyamides. *Antiviral Res*. 2011; 91:177–186. [PubMed: 21669229]
109. Janssen S, Cuvier O, Müller M, Laemmli UK. Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol Cell*. 2000; 6:1013–1024. [PubMed: 11106741]

110. Dickinson LA, Burnett R, Melander C, Edelson BS, Arora PS, Dervan PB, Gottesfeld JM. Arresting cancer proliferation by small-molecule gene regulation. *Chem Biol.* 2004; 11:1583–1594. [PubMed: 15556009]
111. Yang F, Nickols NG, Li BC, Marinov GK, Said JW, Dervan PB. Antitumor activity of a pyrrole-imidazole polyamide. *Proc Natl Acad Sci USA.* 2013; 110:1863–1868. [PubMed: 23319609]
112. Pandian GN, Nakano Y, Sato S, Morinaga H, Bando T, Nagase H, Sugiyama H. A synthetic small molecule for rapid induction of multiple pluripotency genes in mouse embryonic fibroblasts. *Sci Rep.* 2012; 2:544. [PubMed: 22848790]
113. Mapp AK, Ansari AZ, Ptashne M, Dervan PB. Activation of gene expression by small molecule transcription factors. *Proc Natl Acad Sci U S A.* 2000; 97:3930–3935. [PubMed: 10760265]
114. Hauschild KE, Metzler RE, Arndt H-D, Moretti R, Raffaele M, Dervan PB, Ansari AZ. Temperature-sensitive protein–DNA dimerizers. *Proc Natl Acad Sci U S A.* 2005; 102:5008–5013. [PubMed: 15781856]
115. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods.* 2013; 10:1116–1121. [PubMed: 24076762]
116. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* 2013; 21:562–567. [PubMed: 24095303]
117. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154:442–451. [PubMed: 23849981]
118. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* 2013; 10:977–979. [PubMed: 23892898]
119. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013; 152:1173–1183. [PubMed: 23452860]
120. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol.* 2000; 36:244–246. [PubMed: 10760181]
121. Bhaya D, Davison M, Barrangou R. CRISPR–Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet.* 2011; 45:273–297. [PubMed: 22060043]
122. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012; 337:816–821. [PubMed: 22745249]
123. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* 2011; 471:602–607. [PubMed: 21455174]
124. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A.* 2012; 109:E2579–2586. [PubMed: 22949671]
125. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell.* 2013; 50:488–503. [PubMed: 23706818]
126. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat Biotechnol.* 2013; 31:822–826. [PubMed: 23792628]
127. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013; 31:827–832. [PubMed: 23873081]
128. Buskirk AR, Landrigan A, Liu DR. Engineering a ligand-dependent RNA transcriptional activator. *Chem Biol.* 2004; 11:1157–1163. [PubMed: 15324817]

129. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* 2013; 31:833–838. [PubMed: 23907171]
130. Ansari AZ, Mapp AK, Nguyen DH, Dervan PB, Ptashne M. Towards a minimal motif for artificial transcriptional activators. *Chem Biol.* 2001; 8:583–592. [PubMed: 11410377]
131. Wolfe SA, Ramm EI, Pabo CO. Combining structure-based design with phage display to create new Cys₂His₂ zinc finger dimers. *Structure.* 2000; 8:739–750. [PubMed: 10903945]
132. Wolfe SA, Grant RA, Pabo CO. Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry.* 2003; 42:13401–13409. [PubMed: 14621985]
133. Wang BS, Pabo CO. Dimerization of zinc fingers mediated by peptides evolved *in vitro* from random sequences. *Proc Natl Acad Sci U S A.* 1999; 96:9568–9573. [PubMed: 10449733]
134. Wang BS, Grant RA, Pabo CO. Selected peptide extension contacts hydrophobic patch on neighboring zinc finger and mediates dimerization on DNA. *Nat Struct Biol.* 2001; 8:589–593. [PubMed: 11427887]
135. Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J Am Chem Soc.* 2010; 132:6025–6031. [PubMed: 20387835]
136. Koehler AN. A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol.* 2010; 14:331–340. [PubMed: 20395165]
137. Stafford RL, Arndt H-D, Brezinski ML, Ansari AZ, Dervan PB. Minimization of a protein-DNA dimerizer. *J Am Chem Soc.* 2007; 129:2660–2668. [PubMed: 17290996]
138. Xu L, Zerby D, Huang Y, Ji H, Nyanguile OF, de los Angeles JE, Kadan MJ. A versatile framework for the design of ligand-dependent, transgene-specific transcription factors. *Mol Ther.* 2001; 3:262–273. [PubMed: 11237684]
139. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* 2010; 11:636–646. [PubMed: 20717154]
140. Hirai H, Tani T, Katoku-Kikyo N, Kellner S, Karian P, Firpo M, Kikyo N. Radical acceleration of nuclear reprogramming by chromatin remodeling with the transactivation domain of MyoD. *Stem Cells.* 2011; 29:1349–1361. [PubMed: 21732495]
141. Wang Y, Chen J, Hu J-L, Wei X-X, Qin D, Gao J, Zhang L, Jiang J, Li J-S, Liu J, et al. Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep.* 2011; 12:373–378. [PubMed: 21399616]
142. Beerli RR, Segal DJ, Dreier B, Barbas CF. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A.* 1998; 95:14628–14633. [PubMed: 9843940]
143. Tachikawa K, Schröder O, Frey G, Briggs SP, Sera T. Regulation of the endogenous VEGF-A gene by exogenous designed regulatory proteins. *Proc Natl Acad Sci U S A.* 2004; 101:15225–15230. [PubMed: 15475575]
144. Triezenberg SJ, Kingsbury RC, McKnight SL. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* 1988; 2:718–729. [PubMed: 2843425]
145. Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, et al. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem.* 2001; 276:11323–11334. [PubMed: 11145970]
146. Ruben SM, Dillon PJ, Schreck R, Henkel T, Chen CH, Maher M, Baeuerle PA, Rosen CA. Isolation of a rel-related human cDNA that potentially encodes the 65-kD subunit of NF-kappa B. *Science.* 1991; 254:11. [PubMed: 1925549]
147. Pollock R, Issner R, Zoller K, Natesan S, Rivera VM, Clackson T. Delivery of a stringent dimerizer-regulated gene expression system in a single retroviral vector. *Proc Natl Acad Sci U S A.* 2000; 97:13221–13226. [PubMed: 11078518]
148. Giniger E, Ptashne M. Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature.* 1987; 330:670–672. [PubMed: 3317067]

149. Saha S, Ansari AZ, Jarrell KA, Ptashne M, Jarrell KA. RNA sequences that work as transcriptional activating regions. *Nucleic Acids Res.* 2003; 31:1565–1570. [PubMed: 12595565]
150. Minter AR, Brennan BB, Mapp AK. A small molecule transcriptional activation domain. *J Am Chem Soc.* 2004; 126:10504–10505. [PubMed: 15327284]
151. Kwon Y, Arndt H-D, Mao Q, Choi Y, Kawazoe Y, Dervan PB, Uesugi M. Small molecule transcription factor mimic. *J Am Chem Soc.* 2004; 126:15940–15941. [PubMed: 15584709]
152. Xiao X, Yu P, Lim H-S, Sikder D, Kodadek T. Design and synthesis of a cell-permeable synthetic transcription factor mimic. *J Comb Chem.* 2007; 9:592–600. [PubMed: 17530904]
153. Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan W-I, Robson SC, Chung C-W, Hopf C, Savitski MM, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature.* 2011; 478:529–533. [PubMed: 21964340]
154. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I, et al. Selective inhibition of BET bromodomains. *Nature.* 2010; 468:1067–1073. [PubMed: 20871596]
155. Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ. Krüppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A.* 1994; 91:4509–4513. [PubMed: 8183939]
156. Beug H, Müllner EW, Hayman MJ. Insights into erythroid differentiation obtained from studies on avian erythroblastosis virus. *Curr Opin Cell Biol.* 1994; 6:816–824. [PubMed: 7880528]
157. Damm K, Evans RM. Identification of a domain required for oncogenic activity and transcriptional suppression by v-erbA and thyroid-hormone receptor alpha. *Proc Natl Acad Sci U S A.* 1993; 90:10668–10672. [PubMed: 7902566]
158. Snowden AW, Zhang L, Urnov F, Dent C, Jouvenot Y, Zhong X, Rebar EJ, Jamieson AC, Zhang HS, Tan S, et al. Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer Res.* 2003; 63:8968–8976. [PubMed: 14695215]
159. Ayer DE, Laherty CD, Lawrence QA, Armstrong AP, Eisenman RN. Mad proteins contain a dominant transcription repression domain. *Mol Cell Biol.* 1996; 16:5772–5781. [PubMed: 8816491]
160. Sgouras DN, Athanasiou MA, Beal GJ, Fisher RJ, Blair DG, Mavrothalassitis GJ. ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress ets-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation. *EMBO J.* 1995; 14:4781–4793. [PubMed: 7588608]
161. Hiratsu K, Ohta M, Matsui K, Ohme-Takagi M. The SUPERMAN protein is an active repressor whose carboxy-terminal repression domain is required for the development of normal flowers. *FEBS Lett.* 2002; 514:351–354. [PubMed: 11943180]
162. Stormo GD, Zhao Y. Determining the specificity of protein–DNA interactions. *Nat Rev Genet.* 2010; 11:751–760. [PubMed: 20877328]
163. Magnenat L, Blancafort P, Barbas CF III. In vivo selection of combinatorial libraries and designed affinity maturation of polydactyl zinc finger transcription factors for ICAM-1 provides new insights into gene regulation. *J Mol Biol.* 2004; 341:635–649. [PubMed: 15288776]
164. Lund CV, Popkov M, Magnenat L, Barbas CF. Zinc finger transcription factors designed for bispecific coregulation of ErbB2 and ErbB3 receptors: insights into ErbB receptor biology. *Mol Cell Biol.* 2005; 25:9082–9091. [PubMed: 16199884]
165. Tschulena U, Peterson KR, Gonzalez B, Fedosyuk H, Barbas CF. Positive selection of DNA-protein interactions in mammalian cells through phenotypic coupling with retrovirus production. *Nat Struct Mol Biol.* 2009; 16:1195–1199. [PubMed: 19838191]
166. Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol.* 2011; 29:149–153. [PubMed: 21248753]
167. Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, Joung JK. Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods.* 2013; 10:243–245. [PubMed: 23396285]

168. Zhang Z, Xiang D, Heriyanto F, Gao Y, Qian Z, Wu W-S. Dissecting the roles of *miR-302/367* cluster in cellular reprogramming using TALE-based repressor and TALEN. *Stem Cell Rep.* 2013; 1:218–225.
169. Bartsevich VV, Miller JC, Case CC, Pabo CO. Engineered zinc finger proteins for controlling stem cell fate. *Stem Cells.* 2003; 21:632–637. [PubMed: 14595122]
170. Park K-S, Lee D-K, Lee H, Lee Y, Jang Y-S, Kim YH, Yang H-Y, Lee S-I, Seol W, Kim J-S, Lee S-I. Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol.* 2003; 21:1208–1214. [PubMed: 12960965]
171. Lee D-K, Kim YH, Kim J-S, Seol W. Induction and characterization of taxol-resistance phenotypes with a transiently expressed artificial transcriptional activator library. *Nucleic Acids Res.* 2004; 32:e116. [PubMed: 15304545]
172. Blanford P, Chen EI, Gonzalez B, Bergquist S, Zijlstra A, Guthy D, Brachet A, Brakenhoff RH, Quigley JP, Erdmann D, Barbas CF. Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc Natl Acad Sci U S A.* 2005; 102:11716–11721. [PubMed: 16081541]
173. Kwon R-J, Kim SK, Lee S-I, Hwang S-J, Lee GM, Kim J-S, Seol W. Artificial transcription factors increase production of recombinant antibodies in Chinese hamster ovary cells. *Biotechnol Lett.* 2006; 28:9–15. [PubMed: 16369868]
174. Blanford P, Tschan MP, Bergquist S, Guthy D, Brachet A, Sheeter DA, Torbett BE, Erdmann D, Barbas CF. Modulation of drug resistance by artificial transcription factors. *Mol Cancer Ther.* 2008; 7:688–697. [PubMed: 18319335]
175. Gao X, Yang J, Tsang JCH, Ooi J, Wu D, Liu P. Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. *Stem Cell Rep.* 2013; 1:183–197.
176. Xu GL, Bestor TH. Cytosine methylation targeted to pre-determined sequences. *Nat Genet.* 1997; 17:376–378. [PubMed: 9398832]
177. McNamara AR, Hurd PJ, Smith AEF, Ford KG. Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships. *Nucleic Acids Res.* 2002; 30:3818–3830. [PubMed: 12202767]
178. Carvin CD, Parr RD, Kladde MP. Site-selective *in vivo* targeting of cytosine-5 DNA methylation by zinc-finger proteins. *Nucleic Acids Res.* 2003; 31:6493–6501. [PubMed: 14602907]
179. Minczuk M, Papworth MA, Kolasinska P, Murphy MP, Klug A. Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. *Proc Natl Acad Sci U S A.* 2006; 103:19689–19694. [PubMed: 17170133]
180. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell.* 2013; 154:1370–1379. [PubMed: 23992847]
181. DeKolver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijnsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* 2010; 20:1133–1142. [PubMed: 20508142]
182. Soldner F, Laganière J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell.* 2011; 146:318–331. [PubMed: 21757228]
183. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013; 153:910–918. [PubMed: 23643243]
184. Urnov FD, Miller JC, Lee Y-L, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature.* 2005; 435:646–651. [PubMed: 15806097]
185. Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKolver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol.* 2009; 27:851–857. [PubMed: 19680244]
186. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol.* 2011; 29:731–734. [PubMed: 21738127]

187. Doyon JB, Zeitler B, Cheng J, Cheng AT, Cherone JM, Santiago Y, Lee AH, Vo TD, Doyon Y, Miller JC, et al. Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat Cell Biol.* 2011; 13:331–337. [PubMed: 21297641]
188. Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* 2013; 339:823–826. [PubMed: 23287722]
189. Liu W, Yuan JS, Stewart CN. Advanced genetic tools for plant biotechnology. *Nat Rev Genet.* 2013; 14:781–793. [PubMed: 24105275]
190. Facchini LM, Penn LZ. The molecular role of Myc in growth and transformation: recent discoveries lead to new insights. *FASEB J.* 1998; 12:633–651. [PubMed: 9619443]
191. Prelich G. Gene overexpression: uses, mechanisms, and interpretation. *Genetics.* 2012; 190:841–854. [PubMed: 22419077]
192. Perez-Pinera P, Ousterout DG, Brunger JM, Farin AM, Glass KA, Guilak F, Crawford GE, Hartemink AJ, Gersbach CA. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods.* 2013; 10:239–242. [PubMed: 23377379]
193. Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC. Synthetic zinc finger transcription factor action at an endogenous chromosomal site: activation of the human erythropoietin gene. *J Biol Chem.* 2000; 275:33850–33860. [PubMed: 10913152]
194. Gräslund T, Li X, Magnenat L, Popkov M, Barbas CF. Exploring strategies for the design of artificial transcription factors: targeting sites proximal to known regulatory regions for the induction of γ -globin expression and the treatment of sickle cell disease. *J Biol Chem.* 2005; 280:3707–3714. [PubMed: 15537646]
195. Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. *Cell.* 1984; 39:499–509. [PubMed: 6096007]
196. Bacman SR, Williams SL, Pinto M, Peralta S, Moraes CT. Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat Med.* 2013; 19:1111–1113. [PubMed: 23913125]
197. Folmes CDL, Dzeja PP, Nelson TJ, Terzic A. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell.* 2012; 11:596–606. [PubMed: 23122287]
198. Shyh-Chang N, Zhu H, Yvanka de Soysa T, Shinoda G, Seligson MT, Tsanov KM, Nguyen L, Asara JM, Cantley LC, Daley GQ. Lin28 enhances tissue repair by reprogramming cellular metabolism. *Cell.* 2013; 155:778–792. [PubMed: 24209617]
199. Stadtfeld M, Hochedlinger K. Without a trace? PiggyBac-ing toward pluripotency. *Nat Methods.* 2009; 6:329–330. [PubMed: 19404251]
200. Jia F, Wilson KD, Sun N, Gupta DM, Huang M, Li Z, Panetta NJ, Chen ZY, Robbins RC, Kay MA, et al. A nonviral minicircle vector for deriving human iPS cells. *Nat Methods.* 2010; 7:197–199. [PubMed: 20139967]
201. Yoshioka N, Gros E, Li H-R, Kumar S, Deacon DC, Maron C, Muotri AR, Chi NC, Fu X-D, Yu BD, Dowdy SF. Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell.* 2013; 13:246–254. [PubMed: 23910086]
202. Golding MC, Mann MRW. A bidirectional promoter architecture enhances lentiviral transgenesis in embryonic and extraembryonic stem cells. *Gene Ther.* 2011; 18:817–826. [PubMed: 21390068]
203. Gopalkrishnan RV, Christiansen KA, Goldstein NI, DePinho RA, Fisher PB. Use of the human EF-1 α promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: a study using the tetracycline-inducible system in human cancer cells. *Nucleic Acids Res.* 1999; 27:4775–4782. [PubMed: 10572178]
204. Liu J, Jones KL, Sumer H, Verma PJ. Stable transgene expression in human embryonic stem cells after simple chemical transfection. *Mol Reprod Dev.* 2009; 76:580–586. [PubMed: 19034957]
205. Qin JY, Zhang L, Clift KL, Huler I, Xiang AP, Ren B-Z, Lahn BT. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS ONE.* 2010; 5:e10611. [PubMed: 20485554]
206. Lewandoski M. Conditional control of gene expression in the mouse. *Nat Rev Genet.* 2001; 2:743–755. [PubMed: 11584291]

207. Liang J, McLachlan MJ, Zhao H. Orthogonal control of endogenous gene expression in mammalian cells using synthetic ligands. *Biotechnol Bioeng*. 2013; 110:1419–1429. [PubMed: 23242944]
208. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500:472–476. [PubMed: 23877069]
209. Jamieson AC, Miller JC, Pabo CO. Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov*. 2003; 2:361–368. [PubMed: 12750739]
210. Beltran AS, Sun X, Lizardi PM, Blancafort P. Reprogramming epigenetic silencing: artificial transcription factors synergize with chromatin remodeling drugs to reactivate the tumor suppressor mammary serine protease inhibitor. *Mol Cancer Ther*. 2008; 7:1080–1090. [PubMed: 18483297]
211. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schübeler D. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011; 480:490–495. [PubMed: 22170606]
212. Jouvenot Y, Ginjala V, Zhang L, Liu PQ, Oshimura M, Feinberg AP, Wolffe AP, Ohlsson R, Gregory PD. Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. *Gene Ther*. 2003; 10:513–522. [PubMed: 12621455]
213. Bashor CJ, Horwitz AA, Peisajovich SG, Lim WA. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys*. 2010; 39:515–537. [PubMed: 20192780]
214. Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, Chong BG, Ippolito GC, Fujiwara Y, Ebert BL, Tucker PW, Orkin SH. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science*. 2011; 334:993–996. [PubMed: 21998251]
215. Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, et al. Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat Med*. 2002; 8:1427–1432. [PubMed: 12415262]
216. Stolzenburg S, Rots MG, Beltran AS, Rivenbark AG, Yuan X, Qian H, Strahl BD, Blancafort P. Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. *Nucleic Acids Res*. 2012; 40:6725–6740. [PubMed: 22561374]
217. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer*. 2002; 2:777–785. [PubMed: 12360280]
218. Li Y, Ehrhardt K, Zhang MQ, Bleris L. Assembly and validation of versatile transcription activator-like effector libraries. *Sci Rep*. 2014; 4:4857. [PubMed: 24798576]
219. Ouyang Z, Zhou Q, Wong WH. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2009; 106:21521–21526. [PubMed: 19995984]
220. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009; 10:669–680. [PubMed: 19736561]
221. Lee J, Hirsh AS, Wittner BS, Maeder ML, Singavarapu R, Lang M, Janarthanan S, McDermott U, Yajnik V, Ramaswamy S, Joung JK, Sgroi DC. Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS ONE*. 2011; 6:e21112. [PubMed: 21818254]
222. Park KS, Seol W, Yang H-Y, Lee S-I, Kim SK, Kwon R-J, Kim E-J, Roh Y-H, Seong BL, Kim J-S. Identification and use of zinc finger transcription factors that increase production of recombinant proteins in yeast and mammalian cells. *Biotechnol Prog*. 2005; 21:664–670. [PubMed: 15932240]
223. Meissner A. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol*. 2010; 28:1079–1088. [PubMed: 20944600]
224. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011; 25:2227–2241. [PubMed: 22056668]
225. Händel E-M, Alwin S, Cathomen T. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol Ther*. 2009; 17:104–111. [PubMed: 19002164]

226. Prorocic MM, Wenlong D, Olorunniji FJ, Akopian A, Schloetel JG, Hannigan A, McPherson AL, Stark WM. Zinc-finger recombinase activities in vitro. *Nucleic Acids Res.* 2011; 39:9316–9328. [PubMed: 21849325]
227. Pandian GN, Shinohara K-I, Ohtsuki A, Nakano Y, Masafumi M, Bando T, Nagase H, Yamada Y, Watanabe A, et al. Synthetic small molecules for epigenetic activation of pluripotency genes in mouse embryonic fibroblasts. *Chembiochem.* 2011; 12:2822–2828. [PubMed: 22038863]
228. Jüttermann R, Li E, Jaenisch R. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc Natl Acad Sci U S A.* 1994; 91:11797–11801. [PubMed: 7527544]
229. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010; 11:204–220. [PubMed: 20142834]
230. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol.* 2010; 11:607–620. [PubMed: 20683471]
231. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–947. [PubMed: 24119843]
232. Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A.* 1998; 95:10570–10575. [PubMed: 9724744]
233. O'Driscoll M, Jeggo PA. The role of double-strand break repair: insights from human genetics. *Nat Rev Genet.* 2006; 7:45–54. [PubMed: 16369571]
234. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.* 2010; 79:181–211. [PubMed: 20192759]
235. Pâques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther.* 2007; 7:49–66. [PubMed: 17305528]
236. Le Provost F, Lillico S, Passet B, Young R, Whitelaw B, Vilotte J-L. Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol.* 2010; 28:134–141. [PubMed: 20015561]
237. Collin J, Lako M. Concise review: putting a finger on stem cell biology: zinc finger nuclease-driven targeted genetic editing in human pluripotent stem cells. *Stem Cells.* 2011; 29:1021–1033. [PubMed: 21544904]
238. Rahman SH, Maeder ML, Joung JK, Cathomen T. Zinc-finger nucleases for somatic gene therapy: the next frontier. *Hum Gene Ther.* 2011; 22:925–933. [PubMed: 21631241]
239. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* 2012; 14:49–55. [PubMed: 23169466]
240. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 2013; 31:397–405. [PubMed: 23664777]
241. Mussolino C, Cathomen T. On target? Tracing zinc-finger-nuclease specificity. *Nat Methods.* 2011; 8:725–726. [PubMed: 21878917]
242. Pattanayak V, Ramirez CL, Joung JK, Liu DR. Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. *Nat Methods.* 2011; 8:765–770. [PubMed: 21822273]
243. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppl C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol.* 2011; 29:816–823. [PubMed: 21822255]
244. Shimizu Y, Bhakta MS, Segal DJ. Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg Med Chem Lett.* 2009; 19:3970–3972. [PubMed: 19289279]
245. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 2011; 39:9283–9293. [PubMed: 21813459]
246. Wilkins BJ, Rall NA, Ostwal Y, Kruitwagen T, Hiragami-Hamada K, Winkler M, Barral Y, Fischle W, Neumann H. A cascade of histone modifications induces chromatin condensation in mitosis. *Science.* 2014; 343:77–80. [PubMed: 24385627]
247. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell.* 2013; 154:1380–1389. [PubMed: 23992846]

248. Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*. 2010; 140:678–691. [PubMed: 20211137]
249. Hou Z, Zhang Y, Propson NE, Howden SE, Chu L-F, Sontheimer EJ, Thomson JA. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A*. 2013; 110:15644–15649. [PubMed: 23940360]
250. Lombardo A, Cesana D, Genovese P, Di Stefano B, Provati E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P, et al. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods*. 2011; 8:861–869. [PubMed: 21857672]
251. Patterson M, Chan DN, Ha I, Case D, Cui Y, Handel BV, Mikkola HK, Lowry WE. Defining the nature of human pluripotent stem cell progeny. *Cell Res*. 2012; 22:178–193. [PubMed: 21844894]
252. Garg A, Lohmueller JJ, Silver PA, Armel TZ. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res*. 2012; 40:7584–7595. [PubMed: 22581776]

Abbreviations

5-azaC	5-aza-2'-deoxycytidine
ATF	artificial transcription factor
Cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeats
DBD	DNA-binding domain
ED	effector domain
ES	embryonic stem
Exd	extradenticle
HDAC	histone deacetylase
HDR	homology-directed repair
ID	interaction domain
iPS	induced pluripotent stem
KRAB	Krüppel-associated box
LBD	ligand-binding domain
maspin	mammary serine protease inhibitor
NHEJ	non-homologous end-joining
NLS	nuclear localization signal
PAM	protospacer adjacent motif
PNA	peptide nucleic acid
SAHA	suberoylanilide hydroxamic acid
SEAP	secreted embryonic alkaline phosphatase
sgRNA	single-guide RNA

SNP	single nucleotide polymorphism
TALE	transcription activator-like effector
TALEN	TALE nuclease
TF	transcription factor
ZF	zinc finger
ZFN	zinc finger nuclease

Box 1**ZFs**

The Cys₂His₂ ZF motif is the most abundant family of TFs encoded by mammalian genomes, including human genomes, and they target a wide array of genes [48]. Each unit consists of ~30 amino acids which form two antiparallel β -strands and an α -helix, stabilized via a co-ordinated zinc ion [49]. The ZF peptide sequence is (F/Y)-X-C-X₂₋₅-C-X₃-(F/Y)-X₅- ψ -X₂-H-X₃₋₄-H, where X is any amino acid and ψ is any hydrophobic residue, and each finger binds the major groove of DNA [49]. By convention, the critical residues that make specific DNA base contacts are referred to as positions -1, 3 and 6, where -1 is the residue immediately preceding the α -helix [50]. The crystal structure of Zif268/Egr1, a commonly used scaffold in making ZF-based ATFs and nucleases, also shows that the nucleobase contact occurs between the residue at position 2 on the α -helix and a fourth base on the complementary strand [51]. These critical residues can be changed to target specific ~3 bp sequences [52,53]. A catalogue of ZF protein motifs, which target different 3 bp sequences, has been compiled from comprehensive DNA-binding studies [49,52,54-64], and these ZFs are available commercially (Sigma/Sangamo) and non-commercially (Addgene/Zinc Finger Consortium). The modules that are available target GNN sequences the best; the success in targeting a particular triplet is as follows: GNN>ANN>CNN>TNN [65]. Multiple ZF domains can be linked in tandem by the canonical linker TGEKP to target longer and, therefore, less frequently occurring sequences in the genome [50]. The challenge in creating multi-ZF polydactyl proteins is that adjacent fingers can affect binding specificities in non-obvious ways, making it difficult to accurately predict binding preferences, and designed combinations of modules must therefore be validated by secondary assays such as phage display, one- or two-hybrid assays, or cell-based reporter assays [50,53,66-71]. Longer polydactyl ZF proteins do not necessarily contribute to greater specificity because at any given genomic site, a subset of the fingers could bind shorter off-target sites with nanomolar (10^{-9} M) affinity to increase dwell time and affect regulation at undesired loci [72]. The overall consequence of binding to subsets of triplets with nanomolar affinities remains unknown. For a six-ZF protein to fit optimally in the major groove, a glycine or a flexible glycine-serine-glycine tripeptide is added to the canonical linker between fingers 2 and 3 and the linker between 4 and 5 [73]. This 3 x (two-ZF) assembly binds DNA at picomolar (10^{-12} M) affinities and provides specificity better than other architectures tested thus far [74]. Another approach to target longer sequences has been to engineer an ID for co-operative assembly of three- or four-ZF proteins on 18 or 24 bp sites respectively [75,76] (PDB code 1A1L).

Box 2

TALE

TALEs are TFs utilized by the plant pathogen species *Xanthomonas* to alter gene expression of the plant host [77]. TALEs contain a DBD comprising highly conserved ~34-amino-acid tandem repeats, each of which confers exquisite specificity toward a single DNA base. Crystal structures show that each repeat is made up of two α -helices that pack together into a right-handed superhelical structure around major groove of DNA [78,79]. Naturally occurring TALEs typically have 15.5, 17.5 or 19.5 repeats, where the last repeat is a half repeat, and within each 34-amino-acid repeat residues 12 and 13 determine DNA base preference [77]. Residue 12 (typically His or Asn) stabilizes the conformation of the RVD (repeat variable di-residue) loop, and residue 13 makes specific contact with a DNA base [78,79]. Bioinformatic analysis of 113 naturally occurring TALEs followed by experimental validation of RVDs revealed remarkable base preferences: His-Asp targets C; Asn-Gly targets T; Asn-Ile targets A; and Asn-His targets G [77,80-83]. Helical folds N-terminal to the repeats specify a T base, so the target sequence typically must begin with a thymine [79]. Even when taking into account the truncations and optimizations of a commonly used TALE, AvrBs3, so that it is adapted for mammalian cells [77,84,85], a 523-amino-acid polypeptide would be needed to target 9 bp as opposed to ~90 amino acids for a ZF DBD. TALEs are, therefore, inherently larger proteins as base-by-base DNA targeting is achieved by 34-amino-acid modules, whereas a 30-amino-acid ZF module targets a 3 bp stretch. Thus size may be an important property to consider when designing an artificial DNA-binding factor, especially for overcoming the challenges of cellular delivery. A major advantage of TALEs in custom design is that the RVDs seem to function in a context-independent manner. In other words, neighbouring repeats do not affect each other's base specificities, eliminating the screening process that is still required for ZF proteins [86]. Future systematic studies on off-target effects, minimum number of necessary repeats, backbone architecture and the RVD-base pair recognition code will provide more insight on designing TALEs for mammalian systems (PDB code 3UGM).

Box 3**Polyamides**

Minor groove-binding polyamides are a class of molecules that can be rationally designed to target desired DNA sequences [41]. In Nature, Actinobacteria produce the small molecule antibiotics distamycin and netropsin that bind to AT-rich DNA [87,88]. These antibiotics bind genomes of competing bacteria and disrupt DNA-templated processes such as transcription and DNA replication [87,89]. Dervan [90], Dickerson [91], Lown [92] and Wemmer [93] elegantly engineered this natural product scaffold to expand the sequence space that can be targeted. Of the many structures tested, three different heterocyclic rings, pyrrole (Py), imidazole (Im) and hydroxypyrrole (Hp), are commonly conjugated via amide bonds to form crescent-shaped molecules that bind DNA in the minor groove. Incorporation of a short alkyl chain (γ -aminobutyric acid) turn allows polyamide strands to form antiparallel hairpins [94]. Different pairs of rings recognize specific base pairs: Py/Im targets C–G; Im/Py targets G–C; Py/Hp targets A–T; Hp/Py targets T–A; and Py/Py targets A–T or T–A [95]. The side-by-side pairing of these rings is enabled by the alkyl turn that folds the polyamide into a hairpin structure [96]. This pre-organization increases the affinity as well as specificity of polyamides for their targeted cognate sites by 100-fold [90,96].

To accommodate the molecule, the minor groove widens by up to 4 Å, and the opposite major groove contracts in a reciprocal manner [97]. This alteration of DNA shape allosterically inhibits TFs from binding to DNA on the opposite major groove [98]. Although these molecules act as allosteric inhibitors of TF binding, an example of allosteric stabilization has also been reported [99]. Importantly, unlike natural TFs whose binding can be highly affected by the chromatin environment, polyamides can bind repressed regions while still retaining their affinity and specificity towards the targeted sequence. Minor groove-binding polyamides are able to bind methylated DNA because the methyl groups of CpGs protrude into the major groove on the opposite side of where polyamides bind [46]. Biochemical experiments and a crystal structure have shown that polyamides can bind chromatinized DNA in the minor grooves facing outwards with respect to the nucleosome [47,100,101]. Moreover, polyamides can traverse the cell membrane and traffic to the nucleus to exert their effects on gene expression [102–108]. From the induction of striking homoeotic transformations in fruit flies [109], to the inhibition of tumour growth in mice [110,111] and the triggering of the pluripotency network in mammalian cells [112], engineered polyamides can induce remarkable biological outcomes [113,114] (PDB code 3OMJ).

Box 4**Sequence-targeted nucleases**

Top left-hand panel: CRISPR–Cas proteins, derived from the adaptive immune system in prokaryotes, have recently become popular tools for genome editing and transcriptional regulation [115-119]. CRISPR encode Cas proteins, which were originally surmised to bind DNA [120]. Now we know that these proteins target foreign DNA for cleavage, using crRNA (CRISPR RNA) (blue segment) as a guide [121]. Cas9 from the Type II CRISPR–Cas system has been the most amenable for creating breaks within a 20 nucleotide stretch of genomic DNA complementary to crRNA. The maturation of crRNA depends on the tracrRNA (transactivating RNA) (red segment), the RNA molecule that is partially complementary to crRNA and binds Cas9 [116]. Recent work has shown that crRNA and tracrRNA can be fused as one unit called sgRNA [122]. Since RNA, rather than the Cas protein, guides sequence-specific DNA-targeting, preparing the molecular tools for the CRISPR–Cas9 system is much easier than design specifications of ZFs or TALEs for which the protein side chains determine DNA target specificity. Once loaded with sgRNA, Cas9 binds genomic DNA at a site where sgRNA can form base pairs, and this site is juxtaposed to a PAM (yellow segment), a requirement for Cas9 binding [123-125]. Depending on the species from which Cas9 is derived, the PAM sequence is different, but it is generally defined by two to five nucleotides [115,116]. Consequently, genomic targets are restricted by the presence of a PAM sequence. Mismatches within the 8–12 bp stretch proximal to the PAM sequence are not tolerated, whereas mismatches more distal to the PAM sequence can be tolerated, making it difficult to control off-target effects in genomes where longer sequences would be required to target unique sites [126,127]. DNA cleavage occurs through the RuvC-like and HNH-like nuclease domains of Cas9. The D10A and H840A mutations in Cas9 abolish the endonuclease activity of the RuvC-like and HNH-like nuclease domains respectively [122]. Incorporating both mutations results in a programmable sequence-specific DNA-binding factor that lacks nuclease activity. As such, this double mutant functions as a DBD in the design of artificial TFs [115,117-119]. Alternatively, sgRNA fused to RNA ‘ribo-activator’ modules that recruit the transcriptional machinery or even RNA aptamers that recruit TFs can be implemented to regulate transcription [128,129]. Top right-hand panel: DNA cleavage occurs through dimerization of FokI with ZF or TALE DBDs making specific contacts with either side of the target. Bottom panel: After cleavage, the site can undergo HDR or NHEJ. In HDR, a donor plasmid provides a template to create a knockin or to correct or insert a specific mutation. NHEJ leads to random mutations at the site of cleavage. Cleavage at two sites can generate a knockout of a gene.

Figure 1. Toolbox and modular design

(A) Modularity of DNA-binding factors and small molecules. The DBD makes specific contacts with DNA and identifies preferred sites within the genome. An ID can be added to allow the molecule to partner with other proteins of interest. The ED provides the molecule with function. (B) Applications toward controlling gene networks. Regulatory control: a gene-activating ATF can positively regulate gene expression. Conversely, a gene-repressing ATF could negatively regulate its targets. A chromatin-remodelling enzyme fused to a DBD could reorganize the chromatin structure at specific loci where epigenetic control would have important effects in triggering the transcriptional circuitry for a particular phenotype. Genome engineering: attaching a nuclease domain to a DBD would enable genome editing for the purposes of altering transcriptional regulatory elements. In the Figure, reprogramming of fibroblasts to iPS cells is depicted; however, these versatile tools can be applied to a variety of biological contexts. AD, activation domain; RD, repression domain.

Figure 2. Co-operative assembly of TFs

(A) Eukaryotic TFs bind genomic DNA by co-operative assembly. The factors have a different dissociation constant (K_D) depending on whether they bind as monomers or as a larger complex. As a complex, affinity to DNA increases, and binding towards its target site is highly specific. Pol II, RNA polymerase II; TBP, TATA-box-binding protein. (B) A polyamide attached to the YPWM tetrapeptide allows the small molecule DBD to interact with Exd. (C) This interaction with Exd increases affinity for the cognate site three orders of magnitude [99].

Figure 3. Activation domains

To activate genes, a helical peptide with acidic and hydrophobic residues or a small-molecule mimic bearing amphipathic surfaces can be effective in recruiting general transcriptional factors [27]. The transactivator of the herpes simplex virus, VP16, is one such acidic protein that has been shown to activate transcription across eukaryotes [144]. VP64, a tetrameric repeat of the 11-amino-acid core of the activation region from VP16 (DALDDFDLML) is a widely used activation domain (AD) in mammalian systems [142]. In contexts where VP64 has not been effective, RelA (p65) has been used to recruit the general transcriptional machinery [145]. RelA, a subunit of the NF- κ B (nuclear factor κ B) complex, also contains acidic residues interspersed with hydrophobic residues [146]. Pollock et al. [147] created a chimera of RelA and human HSF1, called S3H, which was discovered to be a stronger activator than RelA by itself. Similarly, eleven repeats of the 5-amino-acid minimal activation region of β -catenin (FDTDL) was shown to be a functional activation domain [143]. The amphipathic nature of activation domains led to the design of an AH (amphipathic helix) that displays a negatively charged face and a hydrophobic face [148]. These same principles were applied to the creation of an RNA-based activation domain as well as the isoxazolidine small molecule activation domains [149,150]. Other synthetic activation domains include the small molecule ligand of Sur-2, a subunit of the mammalian Mediator complex (Med), which led to the use of wrenchnolol as an activation module [151]. Kodadek and colleagues also identified multiple peptoids that bind to CREB (cAMP-response-element-binding protein)-binding protein and function as activation domains in human cells [152]. Rather than develop molecules that attempt to block co-operative assembly, one can tether small molecules to a DBD in order to harness co-operative assembly to associate natural TFs at specific sites across the genome. An activation domain could be developed from small molecules that recruit the transcriptional machinery, TFs, chromatin-modifying enzymes as well as chromatin readers. For example,

this strategy could be used to repurpose ligands that bind nuclear hormone receptors such as 17- β -oestradiol and testosterone, or ligands that target the super-elongation complex such as I-BET151 and JQ1 [153,154]. By fusing a small molecule to a DBD, the small molecule can recruit its protein target to activate a gene of interest. Moreover, tethering these molecules to specific loci across the genome via an engineered DBD would also reduce the off-target actions of ligands.

Figure 4. Repressor domains

If the DBD has a high affinity for its binding site, it can function as a competitive inhibitor by occupying potential genomic binding sites of transcriptional activators or interacting with activation domains and preventing them from recruiting the transcriptional machinery. The more effective approach that does not rely on precise placement of competitors at target sites relies on the use of repression modules that recruit repressor complexes or machinery that place transcriptional silencing marks on histone tails. The most widely used repression domain (RD) is the KRAB domain, the 75-amino-acid region at the N-terminus of KOX1 [155]. Extensive investigations of a leukaemia-inducing avian virus led to the finding of v-ErbA as a constitutive repressor [156]. v-ErbA is a variant of the chicken thyroid hormone receptor α , and its mutations render it incapable of activating transcription [157]. Instead, the mutated v-ErbA represses transcription by recruiting HDAC3 and N-CoR-SMRT in human cells [158]. Furthermore, studies of the Mad family of repressor complexes in yeast led to the discovery of the ID of Mad1 (residues 1–36) which functions as a potent repression domain by recruiting co-repressor Sin3A/B [159]. In a screen for factors that bind the *Ets2* promoter, Sgouras et al. [160] screened for a novel ETS protein, which came from a family of transcriptional activators. To their surprise, they stumbled upon a transcriptional repressor, ERF (Ets2 repressor factor), which consists of a repressor domain capable of conferring its repressive function to heterologous DBDs [160]. Even plant repression domains such as the one from the *Arabidopsis thaliana* protein SUPERMAN [161] can function as a repressor in mammalian cells, but, interestingly, they only do so under hypoxic conditions [143]. (KRAB, v-ErbA and MAD1 under PDB codes 1V65, 3N00 and 1S5Q respectively).

1. Choose an accessible site in the gene regulatory element
2. Design DBD to bind the target sequence
3. Attach appropriate ID and ED
4. Test ATF's binding specificity and affinity
5. Use ATF to engineer cell fate

Figure 5. Designing an ATF for a targeted approach

(1) Choose an accessible site in the gene regulatory element as shown in panel A. The location of the TSS (transcription start site) should be taken into consideration. Within the promoter and enhancer elements, DHSs (DNase I hypersensitive sites) shown in red identify regions of accessible chromatin. The most proximal DHS downstream of the TSS (not shown in this example) can also be used for gene regulation [145]. The binding sites for endogenous TFs, depicted as blue boxes, can be considered for competitive binding effects. DHS and TF binding site information is available for some cell types on ENCODE or modENCODE. Some DHSs are better for targeting than others, but selection has to be evaluated experimentally. (2) Design a DBD to bind target sequence. The length of the sequence as well as the nucleotide content should be considered. The rules for DBD design are unique for each class. (3) Attach appropriate ID and ED. The ID is an optional component for co-operative assembly with other TFs. The ED is important for providing the ATF with function (activation or repression). An NLS should also be included to ensure uptake into the nucleus. (4) Test ATF's binding specificity and affinity. Specificity and affinity can be tested *in vitro* by EMSA (electrophoretic mobility shift assay) or more comprehensively by CSI (cognate site identifier) methods and can be inferred in cells by luciferase assay with the target site cloned upstream of luciferase [162]. (5) Use ATF in cells for experiments. The ATF can be expressed in cells by a delivery method of choice to induce changes in cell fate.

Figure 6. ZF library design

ZF modules designed to target GNN, ANN, CNN and TNN sequences are randomized to create polydactyl ZF proteins (depicted is a six-ZF library). Each grey motif represents a module that can target any given triplet. More modules exist for targeting triplets beginning with purines than those beginning with pyrimidines. An ED, such as an AD (activation domain), is used to regulate genes. The library is tested in cells (depicted as fibroblasts or pluripotent cells) by transduction, and hits are selected by screening for a desired phenotype or activation of a lineage-specific gene. ATFs are recovered by sequencing integrated transgenes. Positive ATFs can then be validated for their cell fate-defining effects.

Table 1

Zinc finger ATF libraries used in mammalian cells

Phenotype or reporter gene screened	Library complexity	Number of zinc fingers	Target enrichment	Effector domain	Number of hits validated	Reference
γ -Globulin	1.7×10^7	5	GNN (16), ANN (15), CNN (15)	VP64 or KRAB	4	[165]
CD58, CDH5, EGF, FUT4, ICAM1, ERBB2, ERBB3, ITGA6, ITGB4 and TNFRSF6	1.0×10^4 and 8.4×10^7	3 and 6	GNN (16), ANN (4), TNN (3)	VP64	4 for CDH5	[68]
ICAM-1	8.4×10^7	6	GNN (16), ANN (4), TNN (3)	VP64	4	[163]
Taxol resistance	8.4×10^7	6	GNN (16), ANN (4), TNN (3)	VP64	1	[172]
Drug resistance	1.0×10^4 and 8.4×10^7	3 and 6	GNN (16), ANN (4), TNN (3)	VP64	3	[174]
VEGF	1.5×10^2	3	GNN (28), ANN (4), other (4)	p65 or VP16	4	[60]
Neuronal morphology, ALP	2.4×10^5	4	GNN (37), ANN (9), TNN (2), other (7)	p65 or KRAB	2	[170]
Taxol resistance	Not described	4	25 from [60]	p65	2	[171]
SEAP	2.4×10^5	3 and 4	GNN (37), ANN (9), TNN (2), other (7)	p65 or KRAB	3	[222]
Production of antibodies against TAG72	2.4×10^5	4	GNN (37), ANN (9), TNN (2), other (7)	VP16 or KRAB	4	[173]
Fulvestrant resistance	4.0×10^5	4	GNN (14), ANN (15), CNN (15)	p65	6	[221]